Extending the Reach of Gravitational Wave Detectors and Probing the Isotropic Stochastic Background

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Richard Gary Ormiston

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Vuk Mandic

August, 2021

Acknowledgements

The journey to my doctorate has been, without a doubt, the most physically and mentally taxing experience of my life. Each problem had a weird and wild way of hijacking my brain and begging to be solved. This process was made much more tractable, and dare I say it, downright enjoyable by a number of truly wonderful humans that have my eternal gratitude. To these people, I'm not sure how I can ever repay you, but I will spend a lifetime trying to do so.

To my mentor, Vuk. You responded to a cold email from me and graciously invited me into the group changing my life immeasurably for the better. I have been inspired every day for the last half decade by your creativity, ambition, intelligence, leadership and kindness. You are the embodiment of the ideal scientist and colleague. I cannot understate the privilege it has been to work with you.

To Andrew Matas, my friend, transient roommate, colleague and mentor. You may be the most genuine person on the planet. You have rescued me from levels of confusion so deep that Dante would have given up hope. You are brilliant beyond comprehension and magnificently kindhearted. Our time working together and our brief period as roommates are some of my most cherished memories.

To Pat Meyers, the fastest editor in the universe. You were the first person I met in the LIGO group and you took me on a tour through STAMP-PEM. With dozens of windows open, a beautiful terminal screen and the vim landscape thoroughly ingrained into muscle memory, you blazed through pages at light speed and I was absolutely mesmerized. Instead of making me feel small, you created a path for me, fielded far too many questions from me with nothing but kindness, and inspired me to improve. I owe so much to you and I will be forever grateful for having the good fortune to have crossed paths with you.

To Sharan, my office mate, travel partner and friend. It's almost funny to think about when we met in the lab in the basement of PAN and you were staring at a green and black screen trying to figure out Condor with Tanner's notes and I was still trying to figure out how to log in. From there, we traveled to Caltech, Lake Geneva, Milwaukee and Maastricht. The Netherlands was my only trip out of the country and I'm so happy that I got the chance to make the trip with someone as fun and damn smart as you.

To Michael Coughlin, my friend, advisor, and machine learning partner. There are literally legends about you floating around LIGO and for good reasons. I don't understand how any person can be so dangerously efficient. You have shown me what excellence is every day since I've known you; I'm truly in awe of your talent and perpetually positive attitude. We started a machine learning project some time ago and it knocked my life into another universe. I have learned so much from you and I have nothing but love and admiration for you.

To Andrew Lewis, my best friend of more than 20 years, my partner in crime, my therapist, my brother. Dude... woah. We started off with LAO off-road tricycles and now we're here. I would not have wanted to travel through life with anyone else. You have been *the* major driving force in my life. Anything good that I have accomplished literally could not have been without you. Through thick and thin, giant three egg omelettes at Manny's, kayaking out on the ice, living in seedy apartments, attacking my car with a loaf of bread, picking you up from your first year at UofC, watching you finish law school, we're still the same lovable idiots that we were in fifth grade. Here's to the next 20! I love you, man.

To Melanie, the love of my life. We met right before I started this journey and by some miracle you have stuck by my side without flinching. You have filled my life with love and become the welcomed centerpiece of my future and for that reason I'm perpetually, incalculably, irretrievably ecstatic. I thought Minnesota would be another pit stop, but you turned this place into a home. You've already accomplished so much and your ambition is second to none. I'm inspired by your drive, your wit, your beauty, and your effort to make life about friends and family. I could not have done it without you. I'm convinced that it is impossible for there to be a multiverse in which we didn't find each other.

Lastly, but certainly not least-ly, to Mom and Dad. You were right, it sure is a

lot of debt. It would take a Scrooge McDuck-ish pile to fill this hole. It certainly has been a long walk from signing up for the Jimi Hendrix lab to this, hasn't it? Mom, you never stopped believing in me and did everything in the world to help me out whenever I needed it. Dad, you worked hard so I wouldn't have to and your passion for making things has rubbed off and kept me sane and my work as a physicist grounded. I am fortunate beyond words to have the both of you in my life. Thank you for being who you are.

Alright, let's do this.

Dedication

For Mom and Dad, who encouraged curiosity and happiness above all else.

Abstract

Beginning with the first detection of gravitational waves in 2015 by the Advanced Laser Interferometer Gravitational Wave Observatory (aLIGO) a new era of astrophysics emerged. Within 5 years, aLIGO has detected 50 mergers from binary black hole (BBH) and binary neutron star (BNS) systems and kick started the field of multi-messenger astrophysics with the measurement of an electromagnetic counterpart to the BNS merger GW170817. A detection on the horizon for LIGO is that of the stochastic gravitational wave background (SGWB). The detection of such a background would have far reaching consequences in astrophysics and cosmology as these measurements can probe the first fractions of a second after the Big Bang, revealing insights and parameters of proposed and possibly undiscovered cosmological models. Even a SGWB formed by BBH and BNS mergers near to us would provide valuable information about star formation rates, the formation of large scale structure, as well as the populations of these compact objects.

There are two main topics in this dissertation: detector characterization/data quality methods and characterization of the SGWB. The first chapter provides a background into General Relativity and derives import dynamics relevant to LIGO that are heavily used throughout the thesis. Chapter 2 delves into detector characterization and understanding the noise which enters into the detector output data stream. This is accomplished through the development of a coherence calculation package, `STAMP-PEM`, which creates a hash table lookup for quick followup analysis and a user friendly API.

In Chapter 3 I discuss the sky-averaged SGWB, search methods and present the most recent results combined over aLIGO's first three observing runs. Chapter 4 will extend the isotropic SGWB search to Cosmic Explorer sensitivities and attempt a novel solution to subtract the foreground compact binary coalescences (CBCs) in the $f - t$ space. This mock data analysis sets a benchmark for future search methodologies and sensitivities.

Finally in Chapters 5 and 6 I discuss data quality filtering methods. The former will employ a new deep learning architecture known as `DeepClean` to identify and subtract noise couplings of arbitrary order without introducing artifacts or phase misalignment

of the output signal. The final chapter is dedicated to the construction of analytic filters used for removing linear, nonlinear and non-stationary noise. These analytic filters are useful as they are lightweight and can brute force search through the auxiliary channel combinatorics and aid in identifying relevant physical couplings.

Contents

List of Tables

List of Figures

Chapter 1

Gravitational Waves and Interferometric Detectors

On September 14, 2015, the first direct detection of gravitational waves (GWs) was made by the Laser Interferometer Gravitational Wave Observatory (LIGO) as a result of the merger of a binary black hole system over a billion light years away from Earth [1]. Since then, dozens of other GW signals have been detected [2, 3] including a binary neutron star merger with an electromagnetic counterpart which ushered in the field multi-messenger astrophysics [4]. While GWs had been predicted since the publishing of Einstein's theory of General Relativity in 1915 [5], Einstein himself was unsure if experimentalists would ever be able to achieve the required sensitivity to directly observe gravitational waves.

The implications of gravitational wave astronomy are vast. First, because of the weak interaction and therefore quick decoupling of gravitons, we are in principle able to probe the high energy conditions of the very early universe unlike electromagnetic observations which are limited by the opacity of the last scattering surface that occurred around 380,000 years after the Big Bang [6]. Not only can GWs allow us to see further than EM observations, but these ripples in spacetime are an entirely new observable form of radiation giving us a never-before seen window into the universe. Based on these recent observations, new independent measurements of the universe such as the Hubble constant have been made [7].

Although the early detectors, such as resonant bars, were unsuccessful, the extreme regularity of pulsars was used by Weisberg and Taylor [8] to show that the spin-down of the Hulse-Taylor binary pulsar system [9] was in agreement with the energy that would be lost to gravitational radiation. While the proposal for interferometric detectors had been put forth in the 1970s [10], it was not until the early 1990's that LIGO officially broke ground at Hanford, Washington and Livingston, Louisiana - sites that would eventually became the birth place of gravitational wave astronomy.

The remainder of this chapter will be devoted to the theoretical framework of gravitational waves and the detectors which measure their influences. We will begin with a brief review of special and general relativity, the Einstein equations, weak perturbations and cosmology, and solutions of the wave equations in various regimes. We will also discuss the sources of gravitational waves and their implications, as wall as the hardware and noise characterizations that describes the interferometer-based observatories.

1.1 General Relativity

1.1.1 Special Relativity and Invariance

The principle of relativity states that the laws of physics are unchanged in different inertial reference frames [11]. It was shown that a coordinate transformation exists which invariantly characterizes the temporal and spatial dimensions for all observers assuming a flat spacetime. This was possible by defining the invariant interval as

$$ds^2 = -(cdt)^2 + dx^2 + dy^2 + dz^2 \tag{1.1}$$

This is conveniently represented by a contraction of the four-vector elements dx^μ with the Minkowski metric

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{1.2}$$

as

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu \tag{1.3}$$

where the repeated indices are summed over and run from $0-3$, with 0 being the time coordinate. Using this notation one may define the coordinate transformation

$$\Lambda^\alpha_\beta = \frac{\partial x'^\alpha}{\partial x^\beta} \tag{1.4}$$

and therefore we may transform any vector between coordinate systems as follows

$$dx'^\nu = \Lambda^\nu_\mu dx^\mu \tag{1.5}$$

Maxwell's equations predict that the speed of light in a vacuum is finite. Einstein postulated that this speed was independent of the speed of the source and that all physical processes are the same in all inertial reference frames. This revolutionary idea diverges sharply from the well-known Galilean transformations and formalizes new consequences such as time dilation and length contraction. Special Relativity was a massive triumph, however it suffers from the restriction that it only applies to inertial non-accelerating frames. Ten years later, Einstein introduced the modified theory to include accelerations which we now refer to as General Relativity.

1.1.2 Curved Spacetime

While there are many instances in which spacetime may be considered as flat, it is more generally curved and we therefore promote the metric to one which quantifies the curvature of spacetime, namely $g_{\mu\nu}$. Here the Minkowski metric would be the zeroth order term in the expansion of $g_{\mu\nu}$ and as we will see in section 1.2, the first order metric perturbation gives rise to gravitational waves.

In a curved spacetime, the simple derivative is not sufficient as the application of the simple derivative to a rank-2 tensor gives rise to a term which is not itself a tensor. Therefore, we upgrade to a covariant derivative whose action is given by

$$\begin{aligned}\nabla_\mu v^\sigma &= \frac{\partial v^\sigma}{\partial x^\mu} + \frac{\partial^\sigma}{\partial x'^\rho}\frac{\partial^2 x'^\rho}{\partial x^\alpha \partial x^\mu} \\ &= \frac{\partial v^\sigma}{\partial x^\mu} + \frac{1}{2}g^{\delta\sigma}\left(\partial_\alpha g_{\mu\delta} + \partial_\mu g_{\alpha\delta} - \partial_\delta g_{\alpha\mu}\right)v^\alpha \\ &= \frac{\partial v^\sigma}{\partial x^\mu} + \Gamma^\sigma_{\mu\alpha}v^\alpha \end{aligned} \tag{1.6}$$

where $\Gamma^\sigma_{\mu\alpha}$ is the Christoffel symbol (or affine connection). This derivative then preserves the tensor transformations. Transporting vectors on a curved space poses a problem as there is no unique way to move a vector between tangent spaces - the resultant vector depends on the path taken. To deal with this, we introduce the notion of "parallel transport" defined when the covariant derivative of a tensor along a path vanishes, i.e., when equation (1.6) is zero.

We can now parallel transport a vector in a closed loop over a curved space to find the Riemann curvature tensor which encodes the geometry of spacetime

$$\begin{aligned} R_{\mu\nu\rho}{}^\sigma w^\rho &= \left(\nabla_\nu\nabla_\mu - \nabla_\mu\nabla_\nu\right)w^\rho \\ &= \left(\frac{\partial}{\partial x^\nu}\Gamma^\sigma_{\mu\rho} - \frac{\partial}{\partial x^\mu}\Gamma^\sigma_{\nu\rho} + \Gamma^\sigma_{\nu\lambda}\Gamma^\lambda_{\mu\rho} - \Gamma^\sigma_{\mu\lambda}\Gamma^\lambda_{\nu\rho}\right)w^\rho \\ \rightarrow R_{\mu\nu\rho}{}^\sigma &= \frac{\partial}{\partial x^\nu}\Gamma^\sigma_{\mu\rho} - \frac{\partial}{\partial x^\mu}\Gamma^\sigma_{\nu\rho} + \Gamma^\sigma_{\nu\lambda}\Gamma^\lambda_{\mu\rho} - \Gamma^\sigma_{\mu\lambda}\Gamma^\lambda_{\nu\rho} \end{aligned} \tag{1.7}$$

Let us define other useful quantities such as the contraction of the metric with the Riemann tensor known as the Ricci tensor

$$R_{\mu\rho} = R_{\mu\nu\rho\delta}g^{\nu\delta} \tag{1.8}$$

and the contraction of the Ricci tensor known as the Ricci scalar

$$R = R_{\mu\rho}g^{\mu\rho} \tag{1.9}$$

We may arrange these terms into a divergence-free quantity known as the Einstein tensor by invoking the Bianchi identity and arriving at

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \tag{1.10}$$

where $\nabla^{\mu} G_{\mu\nu} = 0$. This quantity depends on the second derivative of the metric and contains all of the information about the curvature of spacetime.

1.1.3 Einstein Field Equations

While Lorenz, Riemann and others had understood the *geometric* consequences of their formalisms, it was the insight of Einstein that applied these principles to space and time. With a divergence-less curvature tensor now at hand, Einstein proposed that this tensor was proportional to the stress energy tensor $T_{\mu\nu}$ giving

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \alpha T_{\mu\nu} \tag{1.11}$$

Since the stress-energy tensor is also divergence-free, this equation forms the relationship between matter, energy and curvature. The coefficient can be found by taking the Newtonian limit and comparing the results with the classical results

$$\nabla^2 \Phi = 4\pi G \rho \tag{1.12}$$

Doing so gives us Einstein's field equation in its final canonical form

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \frac{8\pi G}{c^4} T_{\mu\nu} \tag{1.13}$$

It is worth noting here that the prefactor to the stress-energy tensor is extremely small, of the order 10^{-43} N^{-1}. This means that in order to achieve any meaningful curvature, $T_{\mu\nu}$ must be enormous. In other words, spacetime is very stiff and it would take cataclysmic events to produce measurable GWs. Lastly, equation (1.13) is not unique. The metric is also divergence-free and could be added to the field equation to the curvature side weighted by Λ,

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \tag{1.14}$$

Einstein referred to this additive term as his "biggest blunder," however it turns out

to be nonzero, specifically $\Lambda = (4.24 \pm 0.11) \times 10^{-66}\text{eV}^2$ [12], and is known as the cosmological constant.

1.2 Gravitational Waves

1.2.1 Weak Field Perturbations

The fluctuations of mass-energy distributions can produce gravitational radiation which distort spacetime. While in the near-zone these distortions can be large, by the time the radiation reaches Earth we may safely consider ourselves to be in the weak-field limit where the metric may be approximated as

$$g_{\mu\nu} \simeq \eta_{\mu\nu} + h_{\mu\nu} \quad \text{where } |h_{\mu\nu}| \ll 1 \tag{1.15}$$

If we consider only a small patch of the metric which can be approximated as locally flat, then the Christoffel connection coefficients become zero and we may write the Riemann curvature tensor as

$$R_{\mu\nu\rho\sigma} = -\frac{1}{2}\left(\partial^2_{\mu\sigma} g_{\nu\rho} + \partial^2_{\nu\rho} g_{\mu\sigma} - \partial^2_{\mu\rho} g_{\nu\sigma} - \partial^2_{\nu\sigma} g_{\mu\rho}\right) \tag{1.16}$$

where we have defined the derivative shorthand notation

$$\partial^2_{\mu\nu} \equiv \frac{\partial^2}{\partial x^\mu \partial x^\nu} \tag{1.17}$$

Inserting the equation (1.15) into equation (1.16) simply replaces $g \to h$ since the Minkowski metric is flat and its derivatives vanish. We may contract equation (1.16) with the metric twice to get the Ricci scalar and therefore write the Einstein tensor in the local frame

$$G_{\mu\nu} = \frac{1}{2}\left(\partial^2_{\mu\sigma} h^\sigma_\nu + \partial^2_{\nu\sigma} h^\sigma_\mu - \partial^2_{\mu\nu} h - \eta^{\sigma\rho}\partial^2_{\sigma\rho} h_{\mu\nu} + \eta_{\mu\nu}\eta^{\sigma\rho}\partial^2_{\sigma\rho} h - \eta_{\mu\nu}\partial^2_{\sigma\rho} h^{\sigma\rho}\right) \tag{1.18}$$

where $\eta_{\mu\nu} h^{\mu\nu} = h$ is the contraction giving the trace of h.

It is often convenient to define a trace reversed metric in order to simplify the above expression. If we define

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h \tag{1.19}$$

then by inserting this into the Einstein field equation two terms vanish and we arrive at the weak field approximation

$$G_{\mu\nu} = \partial^2_{\sigma\nu}\bar{h}^\sigma_\mu + \partial^2_{\mu\sigma}\bar{h}^\sigma_\nu - \eta^{\sigma\rho}\partial^2_{\sigma\rho}\bar{h}_{\mu\nu} - \eta_{\mu\nu}\partial^2_{\sigma\rho}\bar{h}^{\sigma\rho} = \frac{16\pi G}{c^4}T_{\mu\nu} \tag{1.20}$$

If we consider for a moment how the trace-reversed metric changes under the small translations $x^\sigma \to x^\sigma + \xi^\sigma$ we find

$$\bar{h}'_{\mu\nu} = \bar{h}_{\mu\nu} - \partial_\mu\xi_\nu - \partial_\nu\xi_\mu + \eta_{\mu\nu}\eta^{\sigma\rho}\partial_\sigma\xi_\rho \tag{1.21}$$

plus $\mathcal{O}(\xi^2)$ terms which may safely be dropped. To linear order in h, the Riemann tensor is unaffected by this translation and therefore this is a symmetry of the linearized theory. Because we are free to pick ξ^σ, we may choose it such that

$$\partial^\mu \bar{h}^{'}_{\mu\nu} = 0 \tag{1.22}$$

which is known as the "harmonic" gauge. Applying this derivative to equation (1.21) we find

$$\frac{\partial \bar{h}'^\sigma_{\ \nu}}{\partial x'^\sigma} = \frac{\partial \bar{h}^\sigma_\nu}{\partial x^\sigma} - \Box\xi_\nu \tag{1.23}$$

and since the d'Alembertian is an invertible operator, we can always find a transformation such that $\Box\xi_\nu = \partial_\mu\bar{h}^\sigma_\nu$ is satisfied.

By plugging this into equation (1.20) and noting that $\eta^{\sigma\rho}\partial^2_{\sigma\rho} = \Box$ we arrive at

$$\Box\bar{h}_{\mu\nu} = -\frac{16\pi G}{c^4}T_{\mu\nu} \tag{1.24}$$

This is easily recognized as a wave equation where the source term is given by the stress-energy tensor.

While the symmetry in the indices of the metric brings the number of degrees of freedom down from 16 to 10, the choice of gauge further reduces this down to six. We are still free to set four more gauge conditions corresponding to the solutions to $\Box\xi^\sigma = 0$. It

is common to pick a solution which sets $\bar{h} = 0$ and to use the other three to set $\bar{h}^{0i} = 0$. If we apply these conditions to equation (1.22), we see that $\dot{\bar{h}}^{00} = 0$ and therefore $\bar{h}^{00}$ is a constant (2Φ) which we may choose to be zero since we are concerned with the time-dependent behavior. These gauge conditions together are called the "transverse traceless gauge" and we will denote this choice by labeling the metric perturbation $h^{TT}_{\mu\nu}$. It is also worth noting that because the metric is now traceless, equation (1.19) implies that $\bar{h}_{\mu\nu} = h_{\mu\nu}$. The totality of these gauge choices leaves just two degrees of freedom which we shall see in section 1.2.2 correspond to the two polarization states of gravitational radiation.

1.2.2 Solutions to the Wave Equations

In the absence of matter and energy, $T_{\mu\nu} = 0$ and equation (1.24) reduces to

$$\Box\bar{h}_{\mu\nu} = 0 = \left(-\frac{1}{c^2}\frac{d}{dt^2} + \nabla^2\right) h_{\mu\nu} \tag{1.25}$$

and we can immediately recognize this as the wave equation in free space. The simplest solution to this equation is plane waves which we may write as

$$h^{TT}_{\mu\nu} = A\hat{e}^{TT}_{\mu\nu} e^{ik_\sigma x^\sigma} \tag{1.26}$$

where k is the wave-vector, A is the wave amplitude and $\hat{e}_{\mu\nu}$ is the polarization tensor. The harmonic gauge condition tells us that

$$\begin{aligned} \partial^\mu h^{TT}_{\mu\nu} = 0 &= ik^\mu h^{TT}_{\mu\nu} \\ &\rightarrow k^\mu \hat{e}^{TT}_{\mu\nu} = 0 \end{aligned} \tag{1.27}$$

which demands that the polarization of the light is transverse to the direction of propagation. We also observe that the wave equation gives $k^\sigma k_\sigma = 0$ which is consistent with wave propagation at the speed of light shown in equation (1.25).

It is worth pointing out that in the transverse-traceless gauge, $h^{TT}_{0\mu} = 0 \rightarrow h^{TT}_{\mu\nu} = h^{TT}_{ij}$. If we assume that the light is propagating in the $\hat{z}$-direction, then the tensor may be reduced to a 2×2 and from equation (1.26) we have

$$A\hat{e}^{TT}_{\mu\nu} = A\hat{e}^{TT}_{ij} = \begin{pmatrix} h_+ & h_\times \\ h_\times & -h_+ \end{pmatrix} \tag{1.28}$$

This is typically split into a set of polarization tensors called $\hat{e}^{+}_{ij}$ and $\hat{e}^{\times}_{ij}$ given by

$$\hat{e}^{+}_{ij} = h_+ \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \hat{e}^{\times}_{ij} = h_\times \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{1.29}$$

Therefore we may write the general solution to equation (1.26) as $h^{TT}_{ij} = (h_+\hat{e}^{+}_{ij} + h_\times\hat{e}^{\times}_{ij})e^{ik_\sigma x^\sigma}$.

The effect of the gravitational waves on a test mass is given by the change to the proper distance between them. Assuming the test masses start at a distance $\Delta x\,\hat{x}$ away from one another, in the presence of a gravitational wave traveling in a perpendicular direction, the separation becomes

$$\sqrt{ds^2} = \Delta x\sqrt{1 + h_+\cos(\omega t)} \approx 1 + \frac{1}{2}h_+\cos(\omega t) \tag{1.30}$$

A similar case will be true in the y-direction. The net effect of this is that a circular ring of test masses would be squeezed into an ellipse along one axis, rebound, and then squeezed along the other as can be seen in figure 1.1

While we solved for $h_{\mu\nu}$ in a vacuum, it is possible to calculate the general solution to equation (1.24)

$$h_{\mu\nu}(t, \mathbf{x}) = \frac{4G}{c^4}\int \frac{\tau_{\mu\nu}(t - \|\mathbf{x} - \mathbf{x}'\|/c, \mathbf{x}')}{\|\mathbf{x} - \mathbf{x}'\|}d^3\mathbf{x}' \tag{1.31}$$

Here, $\tau_{\mu\nu}$ is the *complete* stress-energy tensor which contains $T_{\mu\nu}$ and the $\mathcal{O}(h^2)$ terms, and the primed coordinates indicate an integration over the source. We may recast equation (1.31) in a more useful form by considering the near and far zone solutions. In the far zone ($R_s \ll \lambda_{GW} \ll r$ where R_s is the radius of the source, λ_{GW} is the wavelength and r is our distance from the source) we have $\|\mathbf{x}-\mathbf{x}'\| \simeq r$ which is essentially constant and therefore we now have

$$h_{\mu\nu}(t, r) \simeq \frac{4G}{c^4 r}\int \tau_{\mu\nu}(t - r/c, \mathbf{x}')d^3\mathbf{x}' \tag{1.32}$$

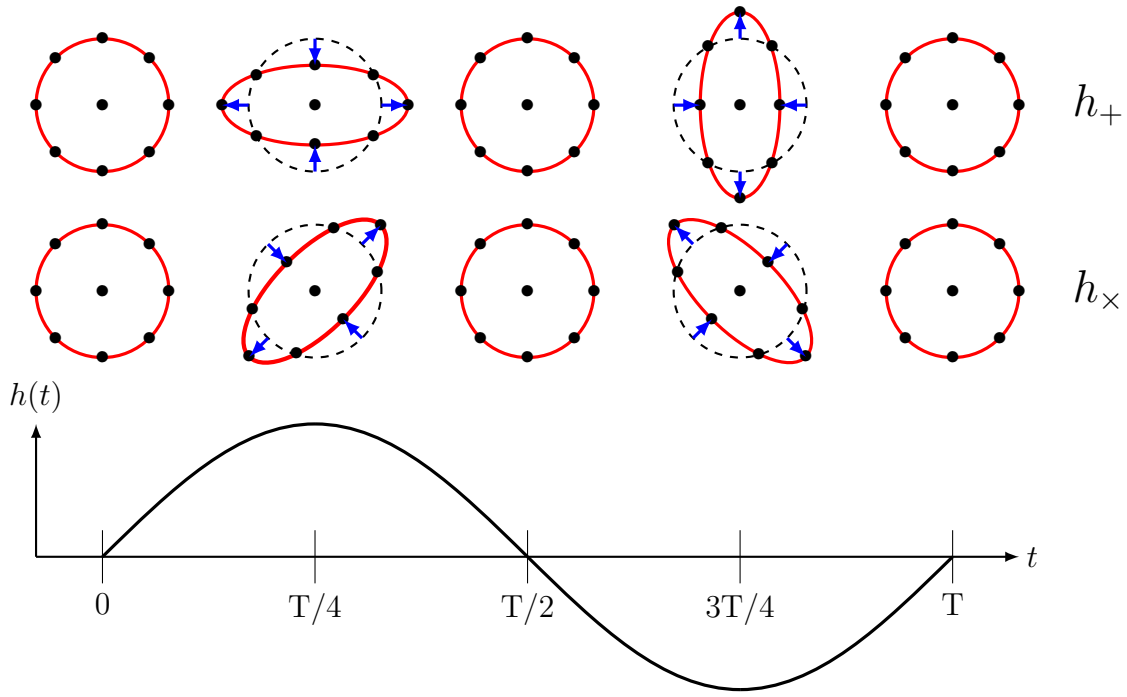

Figure 1.1: The effects of a transverse gravitational wave to a circular ring of test masses during one full period T. The "plus" and "cross" polarizations, h_+ and $h_\times$ respectively, are given their names due to the relative motion of the test masses subjected to a transverse gravitational wave.

Defining the quadrupole tensor to be

$$I_{ij}(t) = \int x'_i x'_j \tau_{00}(t - r/c, \mathbf{x}') d^3\mathbf{x}' \tag{1.33}$$

and using the following identity

$$\tau^{ij} = \frac{1}{2}\partial_0^2 \left(x^i x^j \tau^{00}\right) + \partial_k \left(x^i \tau^{jk} + x^j \tau^{ki}\right) - \frac{1}{2}\partial_{kl}^2 \left(x^i x^j \tau^{kl}\right) \tag{1.34}$$

we may then write the general far zone solution in the more useful form

$$h_{ij}(t, r) \simeq \frac{2G}{c^4 r} \ddot{I}_{ij}(t - r/c) \tag{1.35}$$

This solution may be written in the TT-gauge by applying the symmetric, transverse

projection operator $\hat{P}_{ij} = \delta_{ij} - \hat{n}_i \hat{n}_j$ leading to

$$h_{ij}^{TT}(t,r) = \frac{2G}{c^4 r}\left(\hat{P}_{ik}\ddot{I}^{kl}\hat{P}_{lj} - \frac{1}{2}\hat{P}_{ij}\ddot{I}^{kl}\hat{P}_{kl}\right) \tag{1.36}$$

In the near zone where $R_s \ll r \ll \lambda_{GW}$, the solution is almost the same except that we must use the trace-free quadrupole tensor defined by

$$\mathcal{I}_{ij}^{TT} = \int \left(x_i' x_j' - \frac{1}{3}\delta_{ij} r^2\right) d^3\mathbf{x}' \tag{1.37}$$

The solution is then

$$h_{ij}^{TT}(t,r) \simeq \frac{2G}{c^4 r}\ddot{\mathcal{I}}_{ij}^{TT}(t - r/c) \tag{1.38}$$

For an observer at an inclination angle ι to a rotating binary system with center of mass μ, rotating at a frequency ω and at a distance r away, the polarization amplitudes become

$$\begin{aligned} h_+(t) &= -\frac{4G\mu}{c^2 r}\left(\frac{v}{c}\right)^2 \frac{1+\cos^2(\iota)}{2}\cos(2\omega t) \\ h_\times(t) &= -\frac{4G\mu}{c^2 r}\left(\frac{v}{c}\right)^2 \cos(\iota)\sin(2\omega t) \end{aligned} \tag{1.39}$$

While still measurable by current gravitational wave detectors, typical amplitudes are very small, on the order of $h \sim 10^{-18}$ meters. On the other hand, the luminosities $\mathcal{L}_{GW}$ emitted by an event can be extremely large. To see this, we can calculate the energy radiated away by a source by finding the T_{00} component of the stress-energy tensor. Assuming that the region in which we wish to calculate the GW energy density contains many periods of the wave, we need to perform an integral average over the oscillations by calculating

$$T_{\mu\nu}^{\mathrm{GW}} = -\frac{c^4}{8\pi G}\left\langle R_{\mu\nu}^{(2)} - \frac{1}{2}\eta_{\mu\nu}R^{(2)}\right\rangle \tag{1.40}$$

where the superscripts on the Ricci tensor and scalar denote the order of the expansion to be $\mathcal{O}(h^2)$. After some simplification, this becomes

$$T^{\mathrm{GW}}_{\mu\nu} = \frac{c^4}{32\pi G}\left\langle \partial_\mu h^{ij}_{TT} \partial_\nu h^{TT}_{ij} \right\rangle \tag{1.41}$$

Finally, we can take the $(0,0)$ component of this tensor to get the desired energy density expression,

$$T^{\mathrm{GW}}_{00} = -c\frac{dE}{dAdt} = \frac{c^2}{32\pi G}\left\langle \dot{h}^{ij}_{TT} \dot{h}^{TT}_{ij} \right\rangle \tag{1.42}$$

Integrating over the solid angle subtended by the surface gives

$$-\frac{dE}{dt} = \mathcal{L}_{GW} = \frac{G}{5c^5}\left\langle \dddot{\mathcal{I}}^{TT}_{ij} \dddot{\mathcal{I}}^{ij}_{TT} \right\rangle \tag{1.43}$$

In the case of the gravitational wave event GW150914 which consisted of the the inspiral of two black holes around $30M_\odot$ each, the peak luminosity was $\sim 3.6 \times 10^{56}$ ergs/s or about $200M_\odot c^2$/s [1]. In the merger, $\sim 3M_\odot c^2$ of energy was released in a fraction of a second sending ripples in spacetime across the universe.

1.2.3 Sources and Implications

While any system with a non-vanishing second time derivative of the quadrupole moment produces gravitational waves, in practice these are generally far too small to detect with the current telescopes. Owing to the very small prefactor of equation (1.24), the stress-energy tensor needs to be very large in order to produce a measurable curvature. We will therefore primarily concern ourselves with stars and black holes.

Broadly, sources may be "modeled" or "unmodeled." If a source is modeled then the functional form of its GW signature is understood or well approximated by the current data analysis methods. This is the case of binary black holes (BBH), binary neutron stars (BNS), or black hole−neutron star (BHNS) binaries. These compact objects are of particular interest to ground-based interferometers for a number of reasons: their waveforms are fairly well understood, they pass through the high frequency regime ($\sim$10−1000 Hz) to which the detector is sensitive, they are relatively common, and produce a strain of the order 10^{-21} which is detectable with the current ground-based instruments [13]. The current rate estimate for BNS mergers is $\mathcal{R}_{BNS} = 320^{+490}_{-240}$

$\mathrm{Gpc}^{-3}\mathrm{yr}^{-1}$ and for BBH mergers is $\mathcal{R}_{BBH} = 23.9^{+14.9}_{-8.6}$ $\mathrm{Gpc}^{-3}\mathrm{yr}^{-1}$ [14] and so it is possible in principle for terrestrial GW detectors to measure binary inspiral events daily. By the end of the first half of the third observing run, LIGO-Virgo had accumulated 50 binary merger detections during the course of its three observing runs and these detections are depicted in figure 1.2.

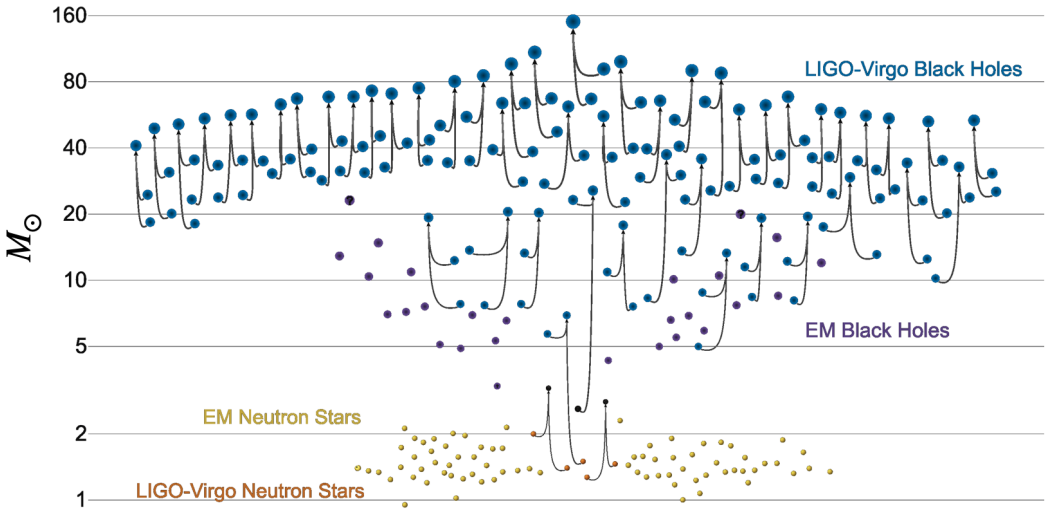

Figure 1.2: Illustration of the stellar graveyard as of O3a (2020) as measured by LIGO and Virgo. There are two binary neutron star events, 47 binary black hole events and one event which is a 23 solar mass black hole with an unknown 2.6 solar mass partner [15]. The grey curves connect the compact objects that formed the merger to the final remnant produced after the collision. Image credit: Frank Elavsky and Aaron Geller of Northwestern University.

As two compact objects orbit each other, they radiate away energy in the form of gravitational waves at twice the orbital frequency of the quadrupole moment. This loss of energy decreases the separation of the compact objects which in turn increases their orbital frequency and hence the radiation given off. The resulting runaway process gives rise to the characteristic "chirp" of compact binary coalescences (CBCs). The frequency evolution of the lead up and merger is analytically described in the quadrupolar, weak-field linearized expansion by the following

$$f_{GW}(t) = \frac{1}{\pi}\left(\frac{5}{256}\frac{1}{(t_c - t)}\right)^{3/8}\left(\frac{GM_c}{c^3}\right)^{-5/8} \tag{1.44}$$

where t_c is the coalescence time of the event and M_c is the chirp mass defined as

$$M_c \equiv \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}} \tag{1.45}$$

The amplitudes of the waveform polarizations also evolve as a function of time (as opposed to the monochromatic solutions given by equation (1.39)) and it may be shown [16] that they are

$$\begin{aligned} h_+(t) &= -\frac{GM_c}{c^2 r}\frac{1+\cos^2(\iota)}{2}\left(\frac{c^3(t_c - t)}{5GM_c}\right)^{-1/4}\cos\left[2\omega t - 2\left(\frac{c^3(t_c - t)}{5GM_c}\right)^{5/8}\right] \\ h_\times(t) &= -\frac{GM_c}{c^2 r}\cos(\iota)\left(\frac{c^3(t_c - t)}{5GM_c}\right)^{-1/4}\sin\left[2\omega t - 2\left(\frac{c^3(t_c - t)}{5GM_c}\right]^{5/8}\right) \end{aligned} \tag{1.46}$$

The behavior of a typical CBC inspiral is shown in figure 1.3 in both the time and frequency domains.

It was found by Fermi et al. [18] that deformations of a neutron star can occur if the energy of the magnetic field surpasses the gravitational potential energy leading to a dynamic instability. Because non-axisymmetric rotating neutron stars have a nonzero second derivative of their quadrupole tensor, these stars will emit gravitational waves. For a rapidly spinning neutron star with ellipticity $\epsilon = (I_{xx} - I_{yy})/I_{zz}$ and rotating at a frequency f, the strain amplitude of GWs given off is [19]

$$h = \frac{16\pi^2 G}{c^4 r} f^2 \epsilon I_{zz} \tag{1.47}$$

Since $\dot{f}_{GW} \sim 10^{-11}$ s^{-2} [20], the frequency is essentially constant over typical LIGO observation times, $\mathcal{O}$(year), and therefore they act as sources for near constant monochromatic sinusoid GWs known as a "continuous wave".

In addition to non-axisymmetric neutron stars, axisymmetric stars can also emit long-lasting gravitational waves by acquiring oscillation modes known as "bar mode instabilities." If a star is collapsing but rotating quickly enough such that the in-fall is

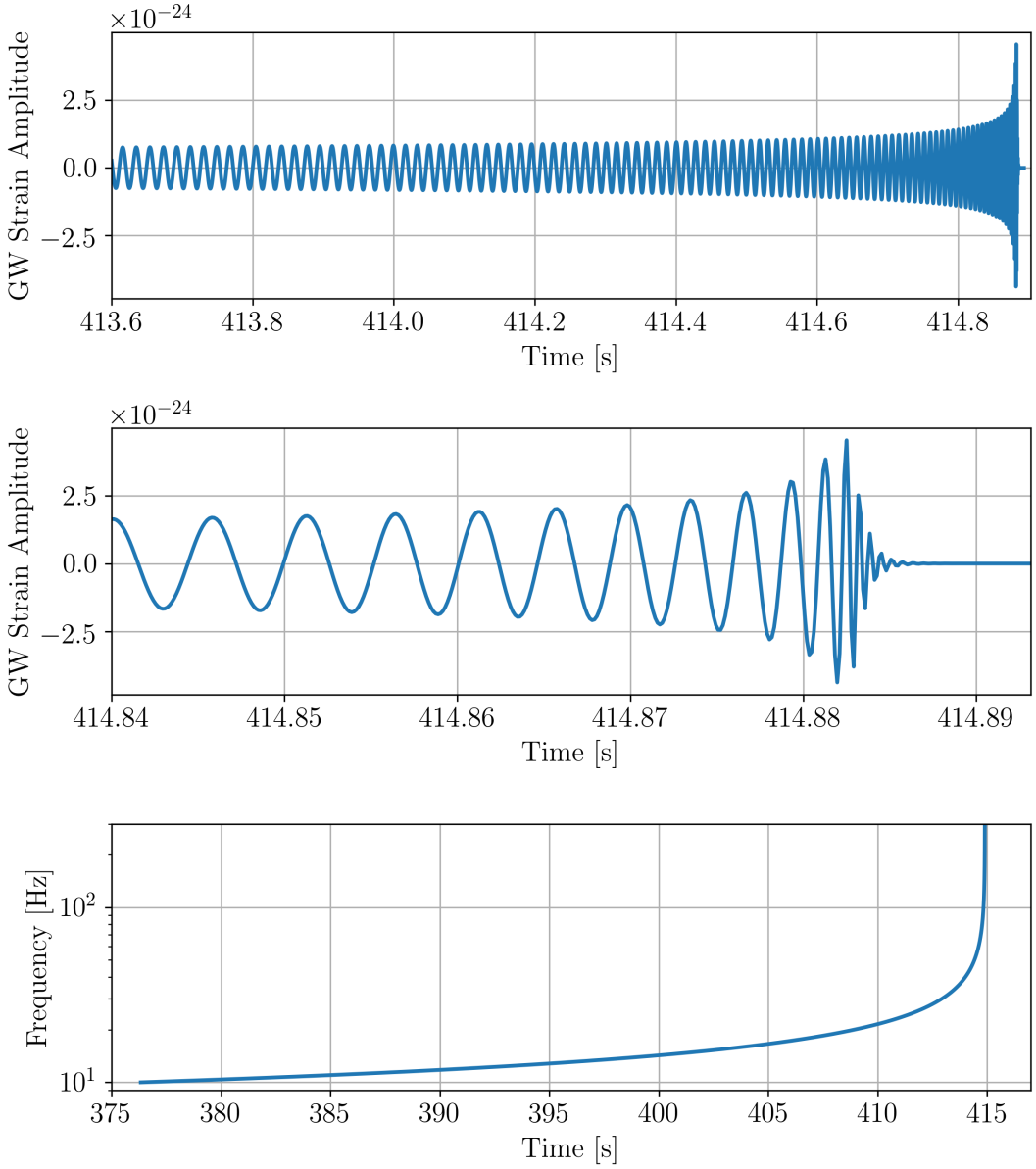

Figure 1.3: A simulation of the inspiral of two $10M_{\odot}$ black holes with no spin and at a luminosity distance of 29 Gpc. Shown is a longer stretch of the waveform evolution (top), a zoomed in version of the final milliseconds showing the merger and ringdown (middle) and the chirp in the frequency evolution. The simulation generating the data uses numerical relativity and is based on an effective-one-body model and was calculated using PyCBC [17].

supported by the centrifugal force, then there can be a production of GWs observed as sharp peaks or "bursts" [21].

In addition to the modeled sources mentioned above, there are also sources with unknown signal models. Core collapse supernovae are one such example. The dynamics depend upon a number of parameters such as the progenitor mass, the rotational frequency and the equation of state which may not be well known. In turn, there is no "signal template" which may be used to identify the signature from events of this kind making them difficult to detect.

Many sources of GWs will be so far away that current detectors will be unable to resolve the signals from these events. The cumulative background from the superposition of gravitational waves from these unresolvable sources is known as the stochastic gravitational wave background (SGWB). Although much more will be said about the SGWB in chapter 3, here we make a few summarizing comments. The background may be from a cosmological origin such as from cosmic strings, primordial black holes, first order phase transitions or inflationary mechanisms and would contain significant information regarding the conditions of the very early universe. Astrophysical sources such as from unresolved events within our galaxy would be distributed anisotropically in a band across the sky corresponding to the galactic plane. Depending upon the original source of the gravitational waves, the frequency spectra will be different and therefore the stochastic background searches use a number of power-law models optimized to detect event-specific contributions. Since the SGWB is characterized by its statistical properties, we use a normalized energy density spectrum to search for the background given by

$$\Omega_{GW}(f) = \frac{f}{\rho_{c,0}} \frac{d\rho_{GW}(f)}{df} \tag{1.48}$$

where $\rho_{c,0}$ is the critical energy density required to close the universe and $\rho_{GW}(f)$ is the energy density spectrum from gravitational waves. As we will show, the background for some SGWBs can be estimated as a power law of the form

$$\Omega_{GW}(f) = \Omega_{\text{ref}} \left(\frac{f}{f_{\text{ref}}} \right)^{\alpha} \tag{1.49}$$

where α is known as the spectral index. Although there is no detection yet, we are able

to place 95% confidence upper limits on Ω_{GW} [22]

$$\begin{aligned} \Omega_{GW} &\leq 5.8 \times 10^{-9} \quad \alpha = 0 \\ \Omega_{GW} &\leq 3.4 \times 10^{-9} \quad \alpha = 2/3 \\ \Omega_{GW} &\leq 3.9 \times 10^{-10} \quad \alpha = 3 \end{aligned} \tag{1.50}$$

1.3 The LIGO Detector Layout

1.3.1 Principles of the LIGO Design

The LIGO detector is fundamentally a Michelson interferometer, although there are a number of modifications which first must be implemented before the detector can reach sufficient strain sensitivity to detect gravitational waves. The actual configuration is shown in figure 1.4 [23].

To see why these alterations are necessary, we note that the strain sensitivity in an interferometer goes like $h \sim \Delta L/L$ where L is the length of the interferometer arms and ΔL is the variation between the length of the arms. For the LIGO Nd:YAG laser operating at a wavelength of 1064 nm inside of 4km arms, the GW metric perturbations would have to be of the order

$$h \sim \frac{\Delta L}{L} \sim \frac{\lambda}{L} \sim \frac{1064 \times 10^{-9}}{4 \times 10^{3}} \sim 2.7 \times 10^{-10} \tag{1.51}$$

which is approximately 11 orders of magnitude away from the required sensitivity. One of the ways to increase the sensitivity of the instrument is to increase the length of the arms. Doing this directly however would be prohibitively expensive and face many challenges such as pumping down a vacuum of that size, and digging the arm cavities partly underground to compensate for the curvature of the earth. Instead, we can increase the *effective* optical path length by introducing additional mirrors into each arm, called input test masses (ITMs), causing the light to bounce back and forth between the ITMs and the mirrors at the end of the interferometer arms, called end test masses (ETMs), many times before exiting the arm cavities. By controlling the reflectivity and transmissivity of the coatings on the test masses forming the Fabry-Pérot cavities,

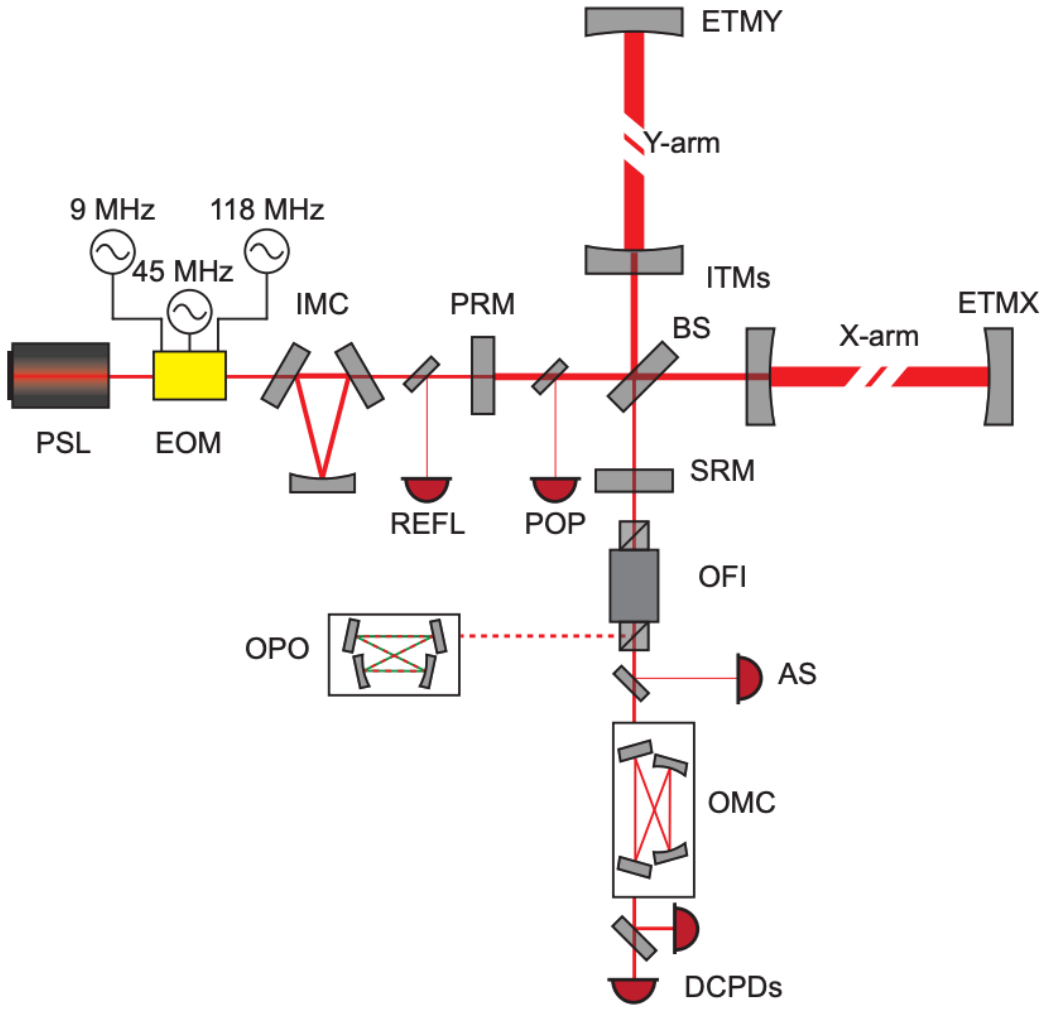

Figure 1.4: The configuration of the optical layout for aLIGO during the third observing run. The initial test masses (ITMs) and the end test masses (ETMs) form the Fabry-Pérot cavities of the interferometer. A power recycling mirror (PRM) feeds light reflected out of the arm towards the laser back into the system to preserve the laser power in the cavities. Some light which exits along the y-direction is recycled by the signal recycling mirror (SRM), however light which is not reflected is blocked from reentering the arms by the output Farady isolator (OFI). The frequency of the laser must remain steady, and so an input mode cleaner (IMC) is used to clean the polarizations, remove unwanted frequencies and beam jitter, as well as to sharpen the beam profile. The output mode cleaner (OMC) removes the carrier light and sidebands (required for Pound-Drever-Hall reflection frequency locking scheme) and the final output strain containing gravitational waves is measured by the direct current output photo diodes (DCPD). Image credit [13].

the average number of bounces of the photons may be adjusted. The effective arm length cannot be arbitrarily high however since for cavity lengths larger than the GWs being measured, the metric would change from the time the light is emitted to the time the light is measured at the anti-symmetric port. Therefore, the wavelength of the GWs determines the maximum optical length of the interferometer and therefore the number of bounces within the resonant cavities. A typical astrophysical gravitational wave measured by LIGO will have a wavelength of order 10^6 m and if we insert this into the denominator of equation (1.51), we find $h \sim 10^{-12}$ which is still much too large.

Before proceeding with the technical hardware additions of the interferometer configuration, consider briefly the fundamental task at hand: Given a laser with wavelength $\lambda \sim 10^{-6}$ m, measure a test mass displacement of $< 10^{-18}$ m. The basic Michelson interferometer measures displacement of $\Delta d \propto \lambda/2$, so it may seem that this detector style would be a lost cause. However, recall that this displacement would be from total destructive interference (a dark fringe) to total constructive interference (a light fringe). We can instead ask *where* on the fringe we are. This amounts to knowing the difference between how many photons arrive at the light fringes and how many actually arrive at the current position. To this end, we can increase the power in the laser and include a power recycling mirror which will feed light exiting along the symmetric port (towards the laser) back into the interferometer keeping the power in the cavities high. The photon arrival is a Poisson process and therefore the fluctuation in the number of photons arriving at any moment (know as "shot noise") goes like $\sqrt{N_\gamma}$ where N_γ is the total number of photons. Fortunately, the strain goes like $\delta N_\gamma / N_\gamma$. For LIGO's 200W laser, the photon arrival rate is

$$N_\gamma \sim \frac{200}{hc}(1064 \times 10^{-9}) \sim 10^{21}\,\mathrm{s}^{-1} \tag{1.52}$$

Including the increase in sensitivity by ∼100x due to the number of bounces of the photons in the Fabry-Pérot cavity [24] and putting together these improvements, we find a strain of

$$h \sim \frac{\Delta L}{L} \sim \frac{1}{\sqrt{N_\gamma}}\frac{\lambda}{\lambda_{GW}} \sim (3 \times 10^{-11})\frac{1064 \times 10^{-9}}{(4 \times 10^3) \times 100} \sim 8.1 \times 10^{-23} \tag{1.53}$$

which is now comparable to the strain of gravitational waves from the binary inspirals that LIGO observes.

Similarly to the limiting returns on the effective optical path length of the laser, we cannot blindly increase the laser power P_{laser} to increase the detector sensitivity. While the shot noise strain scales as $S_{\text{shot}}^{1/2} \propto P_{\text{laser}}^{-1/2}$, the radiation pressure from the photons hitting the end test mass mirrors causes displacements of the mirrors which scale as $S_{\gamma}^{1/2} \propto P_{\text{laser}}^{1/2}$. Therefore we immediately see that there is a trade-off. We will cover this situation and many more in detail in chapter 2 and see how we can minimize the noise across the frequency band of interest.

While aLIGO is fundamentally a simple Michelson interferometer, it is by necessity upgraded to a dual-recycling Fabry-Pérot interferometer. In practice, we still need to contend with many more sources of noise by employing active and passive seismic isolation systems, vacuum chambers, feedback control systems etc. These will be further addressed in chapter 2 as well. The Virgo interferometer in Santo Stefano a Macerata, Italy and the Kamioka Gravitational Wave Detector (KAGRA) in the Gifu Prefecture of Japan operate in much the same way. The noise ASDs of these detectors and of LIGO's "A+" upgrades are shown in figure 1.5

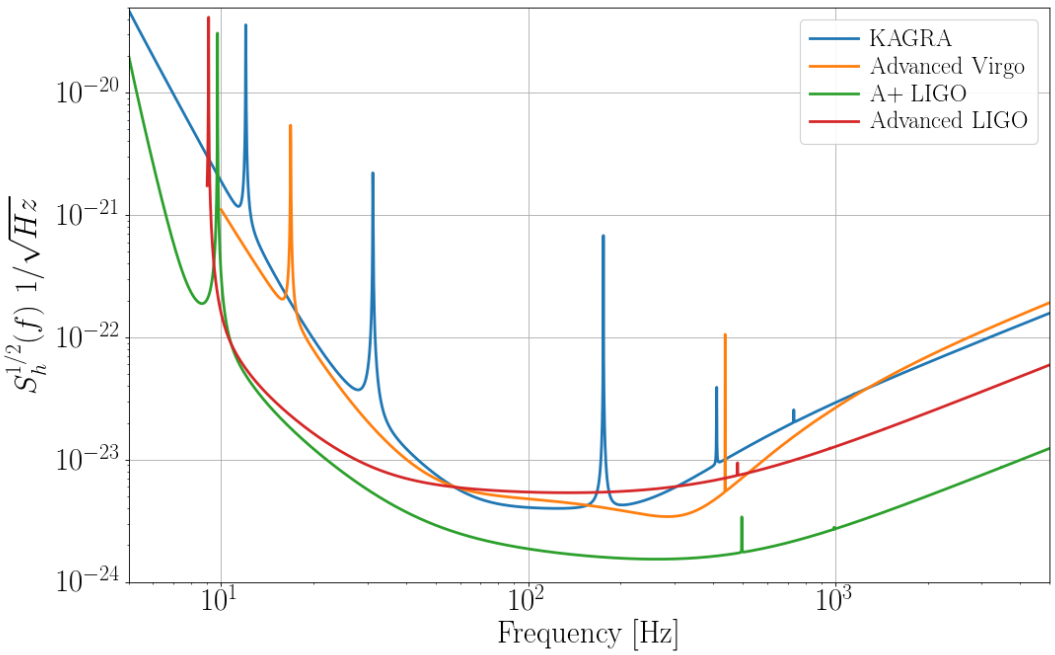

Figure 1.5: Design sensitivities for Advanced LIGO (O4), "A+" LIGO (O5), Advanced Virgo, and KAGRA. Below ∼30 Hz, the seismic noise rises sharply ($\sim f^{-4}$) and limits the sensitivity of ground-based detectors at low frequencies. Above a few hundred Hertz, quantum noise becomes the dominant contribution.

Chapter 2

Noise Sources and Detector Characterization

2.1 Quantification of the LIGO Noise Network

The discovery of gravitational waves has been a tremendous triumph for experimental physics and for hardware and software engineering alike. Confidently making displacement measurements many times smaller than the width of a proton requires an exceptionally sensitive machine as well as a thorough characterization of the detector performance and response to environmental and quantum mechanical fluctuations.

Depending on the origin of the noise, the coupling into the interferometer will vary in strength, stationarity, linearity, mode of coupling and duration. While we aim to understand as much about the noise as possible, not all sources of noise are relevant to all searches. A broadband short duration glitch for example ($\lesssim 1$ s) may be significant to short duration "burst" search pipelines and to the merger of compact binary inspirals such as what happened in the Livingston interferometer during the binary neutron star merger GW170817 [25]. The stochastic gravitational wave background (SGWB) search on the other hand integrates over many months of data and so transients of this type are often less significant[1].

[1]Glitches that alter the variance of the background of adjacent data segments by more than 20% trigger a cut in the stochastic search pipeline which removes the noisy data segment. A large amount of triggers caused nearly half of the data to be removed in O3 and the glitches therefore strongly impacted this search as well. This will be discussed in detail in chapter 3.

Noise sources are observed through the alterations in phase and amplitude to the recombined light on the photodiodes at the "dark" output port of the machine. Since GWs will also produce phase and amplitude shifts observable in the differential arm length measurement readout (DARM), we must be able to separate the terrestrial sensor noise sources from the astrophysical and cosmological signals coupling into the beam. To do this, several thousand auxiliary monitors such as seismometers, magnetometers, microphones and accelerometers are located throughout the detector sites, some of which are shown in figure 2.1. The couplings of the noise sources measured by these auxiliary systems into DARM can be established through hardware and software noise simulations (known as "injections") and subsequently used to identify the origins of noise sources and possible removal of the noise from the strain readout by regression.

An overview of the noise budget for LIGO's third observing run (O3) is shown in figure 2.2. The solid black line labeled "Sum of known noises" represents our understanding of the detector response. Above about 100 Hz, the agreement is quite good whereas at the frequencies below 100 Hz, significant progress remains to be made. The remainder of this chapter will cover the predominant sources of noise within the LIGO interferometer network across various frequency bands, as well as the hardware and software techniques used to understand and combat them.

2.1.1 Injections and Coupling functions

Much of what makes the detection of GWs such a difficult undertaking is the quantification of noise sources ultimately influencing the phase of the beams at the output photodiode. A large effort is made within the LIGO collaboration to create models and corresponding noise mitigation techniques for the various sources of noise [26].

Before noise can be removed or suppressed, the coupling function[2] $C(f)$ of the auxiliary channel to the DARM control loop must be, at least in part, understood. When an auxiliary sensor measures the noise signal in DARM, we say that this data stream is a "witness" channel or simply a "witness". The coupling of the witness channels to DARM is written as

[2]The coupling function is commonly referred to as a "transfer function."

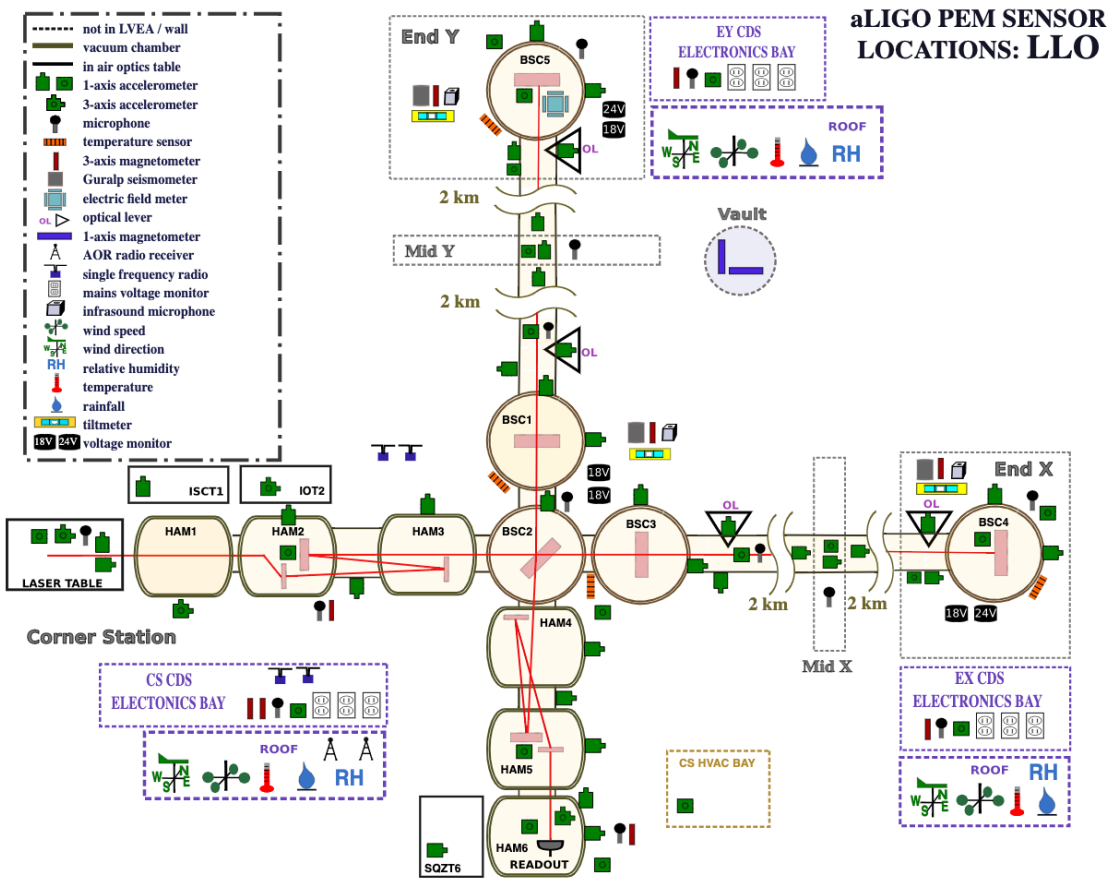

Figure 2.1: Layout of the physical environment monitoring sensors at the LIGO Livingston Observatory during O3. The illustration is by no means exhaustive; tens of thousands of auxiliary channel data streams are recorded to monitor the condition of the detector. Image credit: Maggie Tse, Vincent Roma, Terra Hardwick, Philippe Nguyen (pem.ligo.org/).

$$g(f) = \left(\frac{S^D_{inj}(f) - S^D_{bkg}(f)}{S^W_{inj}(f) - S^W_{bkg}(f)} \right)^{1/2} \tag{2.1}$$

where $S^D_X(f)$ and $S^W_X(f)$ represent the power spectral densities of DARM and witness W during periods of injections (subscript "inj") and normal "quiet" operational conditions (subscript "bkg").

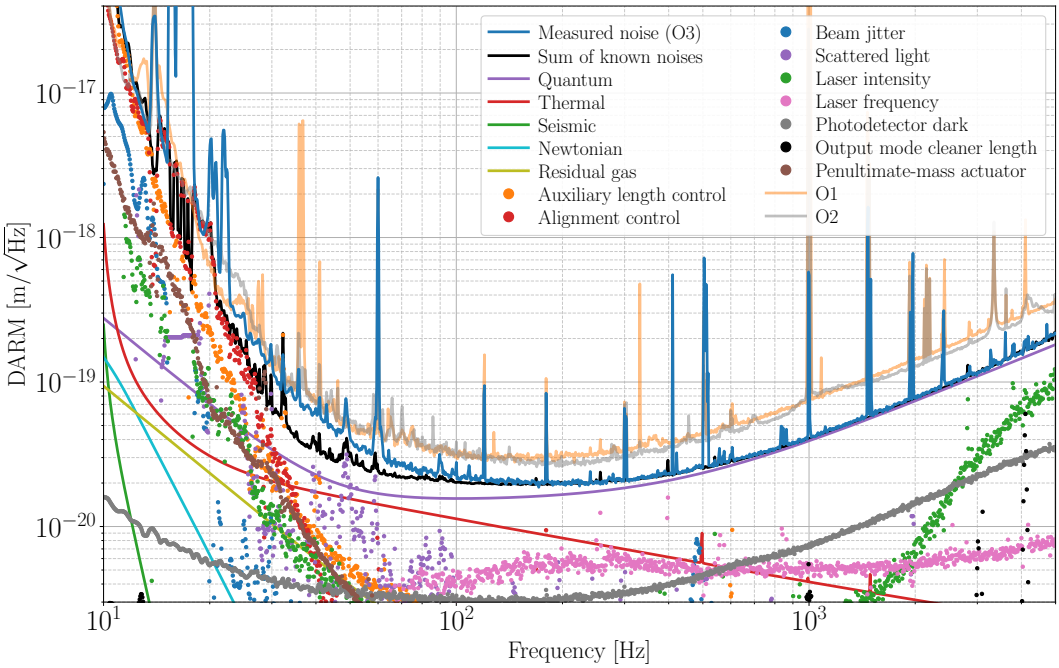

Figure 2.2: Noise budget for the Hanford (LHO) interferometer during O3 [23]. Measured noise sources are denoted with dots whereas calculated models are given by solid lines. For reference, the measured sensitivities for O1 and O2 are also given.

The injections are designed to produce a response in DARM and remain in the sensitive band of the witness sensor without saturation which would cause an overestimate of the coupling. Different techniques are used to measure the coupling of different witness channels: acoustic injection generated by speakers playing a predetermined range of frequencies are used to drive accelerometers (figure 2.3), actuated pistons called "shakers" are used to create seismic influences, large coils are used to produce locally constant magnetic fields etc.

2.1.2 Metrics of Interferometer Performance

Inexorably tied to the characterization of the noise network at LIGO is the question of how to accurately quantify the state of the detector. Noise will always pervade the instrument, but we must develop "figures of merit" to determine if the effect of the noise

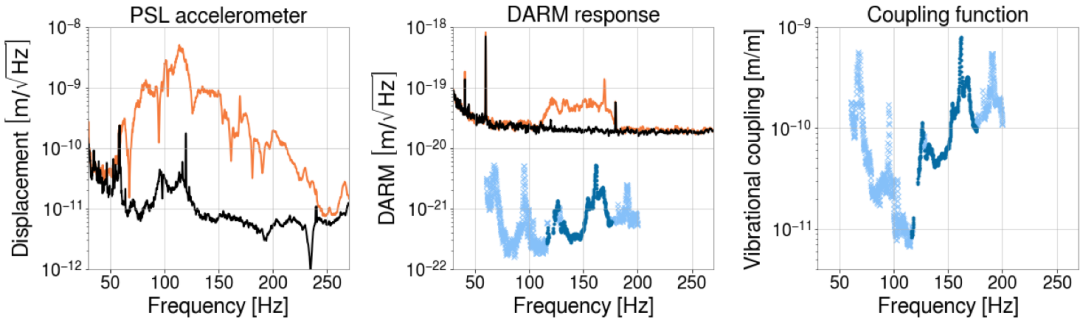

Figure 2.3: *Left*: Amplitude spectral density (ASD) of an acoustic injection used to measure the response of the pre-stabilized laser (PSL) table accelerometer. The orange curve is the accelerometer response during the injection and the black curve is before the injection. *Middle*: The ASD of DARM during (orange) and before (black) the injection. *Right*: The calculated coupling function of the influence measured by the PSL accelerometer to DARM. Image sourced from [27].

is enough to corrupt all or part of the data. Presently, there are a number of standard indicators used by the LIGO collaboration including the BNS inspiral range, duty cycle, and the strain sensitivity curve.

The BNS inspiral range is a relative measure of the observable range of a single detector by quantifying out to what distance a merger of $1.4M_\odot$ neutron stars could be detected with a signal-to-noise ratio (SNR) ≥ 8. Generally, the inspiral range is given in [28] by the formula

$$r^2 = \frac{5\mathcal{M}^{5/3}\theta^2}{96\pi^{4/3}\rho_0^2}\int_0^\infty \frac{1}{f^{7/3}S_n(f)}\,df \tag{2.2}$$

where r is the range, $\theta = 1.77$ and accounts for averaging binaries' sky positions and orientations relative to the detector's antenna pattern [29], ρ_0 is the required minimum SNR, $S_n(f)$ is the one-sided PSD of the noise, and $\mathcal{M}$ is the chirp mass as defined in equation (1.45) and reduces to $1.2M_\odot$ given the assumptions above. We see that the range scales roughly as $r \sim S_n^{-1/2}(f)$ and so increases in the noise floor decrease the sensitivity and observable volume of the detector. As measurements of the BNS inspiral range are taken every minute, even transient noise sources can cause noticeable dips in the range. During the commissioning break between the second and third observing

runs, a number of hardware upgrades were made and the average BNS inspiral range was increased in each LIGO detector and in Virgo as seen in figure 2.4

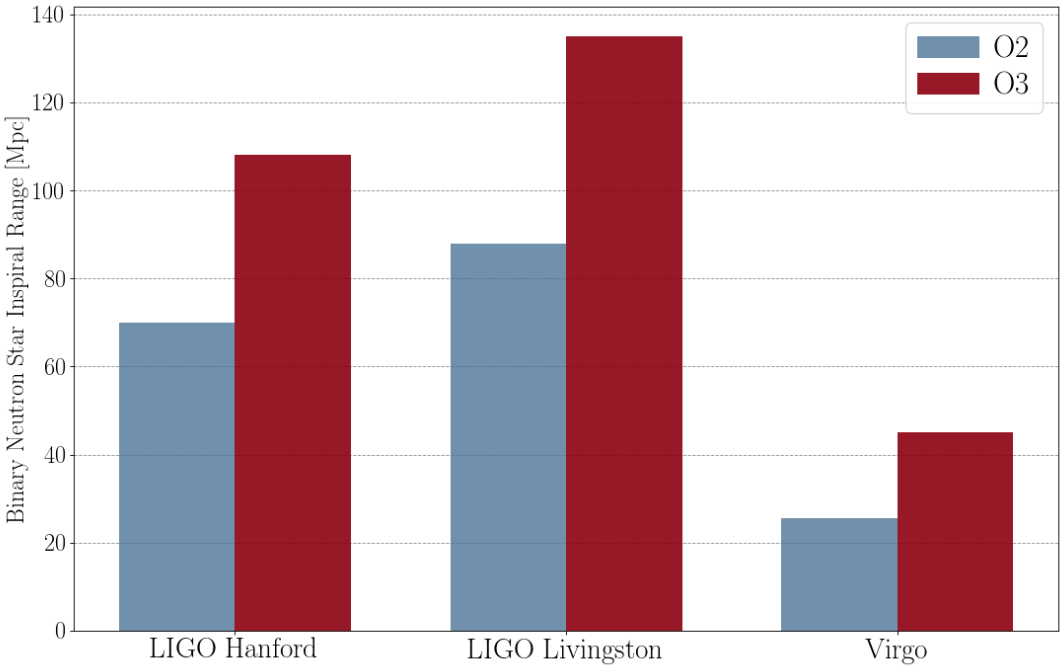

Figure 2.4: The average distance at which two neutron stars with mass $1.4M_{\odot}$ could be detected in a single detector with an SNR ≥ 8. While Virgo is not as sensitive as the LIGO detectors, it is able to detect GWs and aid in sky localization [30]. The Livingston detector suffered from light scattering noise during the beginning of the O3 run, although this was addressed during operation.

2.2 Quantum Noise

While there are a number of noise sources that plague the detector which may be directly monitored, quantum mechanical sources of noise can only be modeled statistically and form a fundamental noise floor to the detector which cannot be removed either through hardware[3] or software techniques. The quantum noise is a result of the small fluctuations

[3]This is not strictly true. Light squeezing is able to reduce quantum noise in the degree of freedom that we care about by approximately a factor of 2 above ∼100 Hz [31]

in the electric field of the laser and may be separated into two categories, "shot noise" and "radiation pressure".

2.2.1 Shot Noise

As mentioned in section 1.3.1, in order for the interferometers to measure displacements on the order $\sim 10^{-18}$ m with a laser of wavelength $\sim 1\,\mu$m, it is imperative to have a large number of photons in the arm cavity. This increases the sensitivity of the photodiodes as to where on the fringe the detector sits. Due to quantum mechanical fluctuations, the number of photons will spontaneously change, therefore making the number of photons at the output port vary even for identical strains. Because the number of photons N_γ is large and the uncertainty is driven by a Poisson process, the number of photons arriving at the output decreases with $\sqrt{N_\gamma}$. The expression for the shot noise strain is given by

$$S_{\text{shot}}^{1/2}(f) = \sqrt{\frac{\hbar\lambda c}{\pi L^2 I_0 G_{prc} G_{arm}^2}}\sqrt{1 + (f/f_p)^2} \tag{2.3}$$

where λ is the laser wavelength, G_{prc} is the gain in the power recycling cavity which returns light from the symmetric port back into the cavity arms, G_{arm} is the gain inside the arm cavities, L is the length of the cavity arms, and f_p is the pole frequency (~ 85 Hz) of the arm cavity. It would consequently seem advantageous to increase the beam intensity (or decrease the wavelength) as much as possible, however this is not so. The number of photons hitting the ITM and ETM also fluctuate and this imbalanced pressure induces a torque on the mirrors causing beam alignment issues and therefore also contributes to the noise. This is known as the "radiation pressure".

2.2.2 Radiation Pressure

While the shot noise can be improved by increasing the beam power, this also increases number of photons hitting the test masses. Since there will be fluctuations in the number of photons hitting the mirrors, the beam will impart slightly different forces onto different parts of the mirror. This imbalance can (and does) cause the mirrors to oscillate and move the beam spot out of alignment with the photodiodes. This noise is described by the following

$$S_{\rm rad}^{1/2}(f) = \sqrt{\frac{2\hbar I_0 G_{prc} G_{arm}^2 C^2(f)}{\pi^2 m^2 L^2 c f^4}} \tag{2.4}$$

where $C(f)$ is the "sensing function" which relates the light intensity at the anti-symmetric port with the GW strain, m is the mass of the test mass, and L is the length of the cavity arms. Comparing equation (2.3) with equation (2.4), it is evident that minimizing radiation pressure noise through the beam intensity or wavelength would result in increasing the shot noise. We can however increase the mirror's mass to reduce the effect of radiation pressure. The upgrades from Initial LIGO to Advanced LIGO included increasing the mirrors from 11kg to 40kg for precisely this reason. The net effect of a larger test mass is to push the the radiation pressure noise to low frequencies where it does not affect the most sensitive frequency band, or "bucket" of LIGO.

What is usually done is to minimize the sum of the quantum noises with respect to the laser intensity,

$$\frac{d}{dI_0}\left(S_{\rm shot}^{1/2}(f) + S_{\rm rad}^{1/2}(f)\right) = 0 \tag{2.5}$$

This results in a minimum total quantum noise, known as the "standard quantum limit",

$$S_{\rm quantum}^{1/2}(f) = \sqrt{\frac{8\hbar}{(2\pi f)^2 m L^2}} \tag{2.6}$$

This is the minimum noise achievable at *any* given frequency. Achieving this lower limit requires an optimization at each frequencies and it is therefore not possible to obtain this limit at multiple frequencies simultaneously.

2.3 Displacement Noise

A general designation for sources of noise which cause a change in the differential arm cavity length readout is "displacement noise." While there are many contributions to this noise, here we concern ourselves only with the sources which dominate a particular frequency band that falls within the $\sim$10$-$100 Hz range.

2.3.1 Seismic Noise

The interferometers have been designed with particular dimensions intended to produce destructive interference of the returning beams at the output port. This interference is of course sensitive to the length of the arm cavities and the orientations of the mirrors relative to the beam, beam splitter and a number of other optical components. The ground motion near the detector sites caused from Earth's internal seismic activity, anthropogenic sources and the regularity of ocean waves hitting the shore therefore poses a problem.

The strain $S_g^{1/2}(f)$ due to the relative ground motion can be roughly described as a broken power law [32]

$$S_g^{1/2}(f) \simeq \begin{cases} 10^{-12}\ \frac{1}{\sqrt{\text{Hz}}} & \text{for } f < 10 \\ 10^{-12} \left(\frac{10}{f}\right)^2\ \frac{1}{\sqrt{\text{Hz}}} & \text{for } f \geq 10 \end{cases} \tag{2.7}$$

At 10 Hz, the ground motion itself is at $\sim 10^{-9}\,\text{m}/\sqrt{\text{Hz}}$, but the length of the arms gives an extra three orders of magnitude suppression. This is clearly far above the required strain of order 10^{-22} within the LIGO frequency band. Two different methods are used in order to push the seismic noise floor down and make GW astronomy possible, passive and active isolation.

The passive isolation takes the form of a simple pendulum. Consider the oscillations of a pendulum bob under the influence of time varying displacement and without damping

$$\ddot{x}(t) + \omega_0^2\big(x(t) - A(t)\big) = 0 \tag{2.8}$$

where $\omega_0 = \sqrt{g/L}$ is the resonant frequency of the pendulum and $A(t)$ is the time dependent driving force. In the Fourier domain this becomes

$$(-2\pi i f)^2 \tilde{x}(f) + (2\pi f_0)^2 \tilde{x}(f) = (2\pi f_0)^2 \tilde{A}(f)$$

$$\begin{aligned} \rightarrow \tilde{x}(f) &= \frac{1}{1-(f/f_0)^2}\tilde{A}(f) \\ &= T(f)\tilde{A}(f) \end{aligned} \tag{2.9}$$

where $T(f)$ is the system's transfer function. We can see now that the motion of $\tilde{A}(f)$ on $\tilde{x}(f)$ is suppressed by a factor of f^{-2} when $f > f_0$. This is implemented at LIGO, but it is not enough. The fundamental resonant frequency of the pendulum is $f_0 \simeq 0.76$ Hz [24] and therefore at 100 Hz we have a strain of

$$S_g^{1/2}(f) \sim 10^{-12}\left(\frac{10}{100}\right)^2\left(\frac{0.76}{100}\right)^2 \sim 5.8\times 10^{-19}\,1/\sqrt{\text{Hz}} \tag{2.10}$$

At this level, even at 100 Hz the GW signals will be a few orders of magnitude below the seismic noise and will not be observed. One solution is to chain multiple pendula together and therefore multiply the transfer functions together increasing the total suppression factor. This technique is implemented at LIGO (figure 2.5) where the test masses are suspended in a quadruple pendulum resulting a seismic strain at and above 10 Hz of

$$S_g^{1/2}(f) \sim 10^{-12}\left(\frac{10}{f}\right)^2\left[\frac{1}{1-(f/0.76)^2}\right]^4\frac{1}{\sqrt{\text{Hz}}} \tag{2.11}$$

The result is that now at 10 Hz, $S_g^{1/2}(f=10) \sim 10^{-21}$ Hz$^{-1/2}$ which is generally sufficient. The beamsplitter, input mode cleaner (IMC) and recycling cavities are hung from a triple pendulum. The output mode cleaner (OMC) and end test mass transmission monitor are hung from a double pendulum.

The quadrupole pendulum is not the only source of seismic isolation; aLIGO has added additional active feedback isolation systems known as "Internal Seismic Isolation" (ISI). The ISI receives seismic data from a number of sensors and effectively inverts signals which fall below ∼1.5 Hz. These inverted control signals are sent into a linear feedback control servo which in turn drives electrostatic actuators on the ISI platform and down the reaction chain [34]. The result is that the incoming displacement noise is counteracted by the active isolation. The noise contributions which make it beyond the ISI are then coupled into the quad pendulum and further suppressed by $\sim f^{-8}$ as described above in equation (2.11).

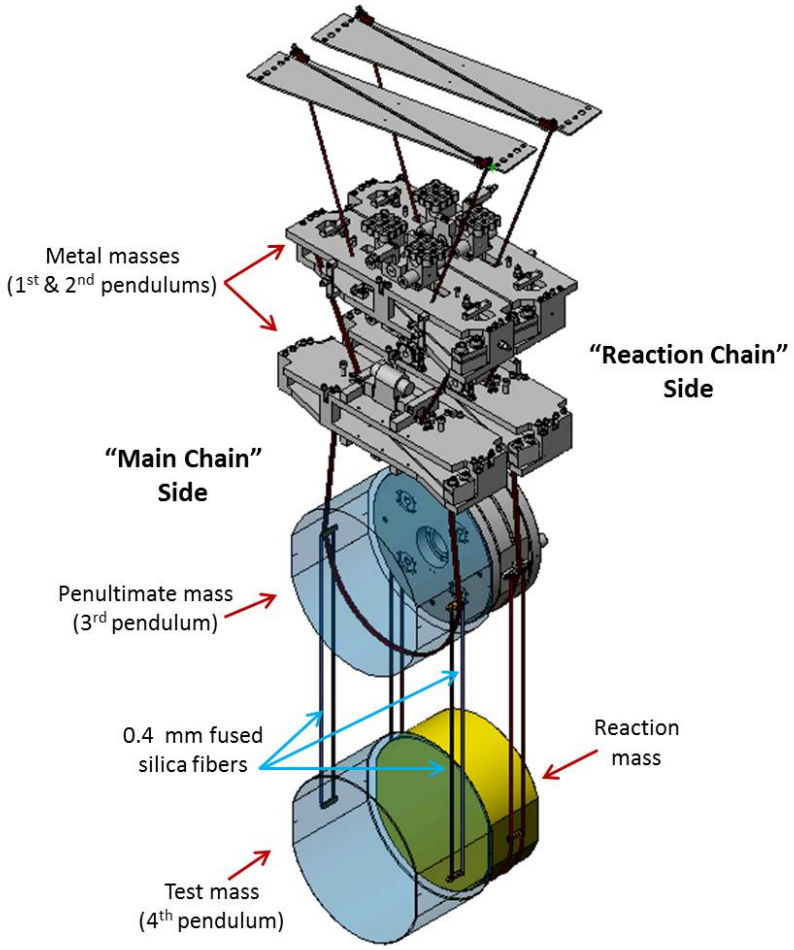

Figure 2.5: Representation of LIGO's quadruple pendulum suspension system for passive seismic isolation [33]. The "Main Chain" side of the mirror faces the laser. The weight of the test mass mirror also aids in the dampening of displacement noises.

2.3.2 Thermal Noise

Thermal noise arises in the system in a few ways: fluctuations from thermal gradients in fused silica suspension wires, Brownian motion of the coating on the test mass surfaces, and laser heating.

The suspension noise becomes the dominant source of noise below ~ 10 Hz if the seismic noise has been adequately filtered. This noise source drops off as $f^{5/2}$ however, and

is not a significant limitation of the detector's sensitivity above 10 Hz with the exception of the excitation of the violin modes at ∼500 Hz and higher harmonics. Attempts are made nonetheless to reduce the thermal suspension noise. The fluctuation-dissipation theorem states that fluctuations cause dissipation, or more directly we can write the power spectral density as

$$S_x(f) = \frac{k_B T}{\pi^2 f^2} \Re \left[\mathcal{Z}^{-1}(f) \right] \tag{2.12}$$

where T is the temperature, k_B is Boltzmann's constant and $\mathcal{Z}$ is the impedance of the system. Because the thermal suspension strain will go like $S_x^{1/2} \sim \sqrt{T}$ very low temperatures are needed to create improvements comparable to other methods, although this is still a realistic and actively pursued avenue for future detector upgrades. Current developments however favor mechanical enhancements, such as to the mirror coating and suspension wires, which reduce the dissipation to the longitudinal modes of the mirror that would negatively affect the output strain measurements.

The coating thermal noise (CTN) on the other hand is the second most dominant contribution to the noise near 100 Hz, second only to the quantum noise. From figure 2.6 it can be seen that the amplitude spectral density for the CTN spectrum is relatively flat and can be shown to follow [35]

$$N_{CTN}(f) = (7.5 \pm 0.1) \times 10^{-21} \left(\frac{100 \text{ Hz}}{f} \right)^{0.45 \pm 0.02} \frac{\text{m}}{\sqrt{\text{Hz}}} \tag{2.13}$$

In addition to the CTN and thermal suspension noise, small thermo-elastic distortions in the ITM lenses from the beam power can degrade the shape of the reflected TEM modes and therefore the sensitivity. To combat this, CO_2 lasers are aimed at the ITMs to provide controlled heating and maintain the lens shape, and are adjusted to maximize the cavity gain and sensitivity [36].

2.3.3 Electrical Noise

In the United States, the power lines supply alternating current at a frequency of 60 Hz. The side effect of this is that the electronics used to not only measure the detector's status, but to actuate the mirrors as part of the linear feedback loop, fluctuate at this frequency as they draw power. This causes the mirrors to oscillate at 60 Hz and the

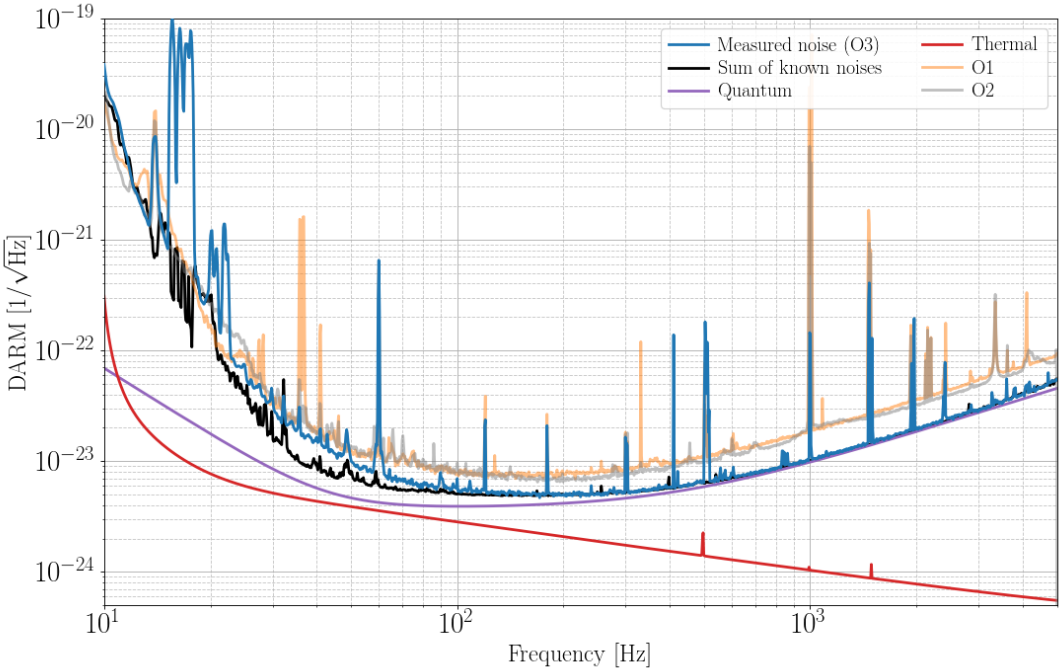

Figure 2.6: The quantum and thermal noise contributions to the strain of the LHO spectrum during O3. At frequencies near 100 Hz the thermal noise, dominated by the coating thermal noise, contributes significantly to the noise floor of the detector. For comparison, the aLIGO O1 and O2 curves are also included to show the progress in reducing the detector strain suggesting that coating properties will become increasingly important. This figure has been reproduced with permission from G. Mendell and the writing team of [23].

higher harmonics. The power build-up at these frequencies, visible in figure 2.7, is many orders of magnitude above the surrounding noise floor and very narrow. Fortunately, magnetometers and well as voltage mains monitors are sensitive to these voltage changes as well and can be used to remove the excess electrical power [37, 38].

In addition to the noise introduced due to the 60 Hz AC current supply, electronics which draw power periodically can lead to increased power at the fundamental frequency and the higher order resonances. This happened during O1 at the US interferometers when LEDs blinking at 1Hz intervals caused narrow spectral artifacts to appear at every integer frequency across LIGO's observing band [39]. While the frequencies of the power

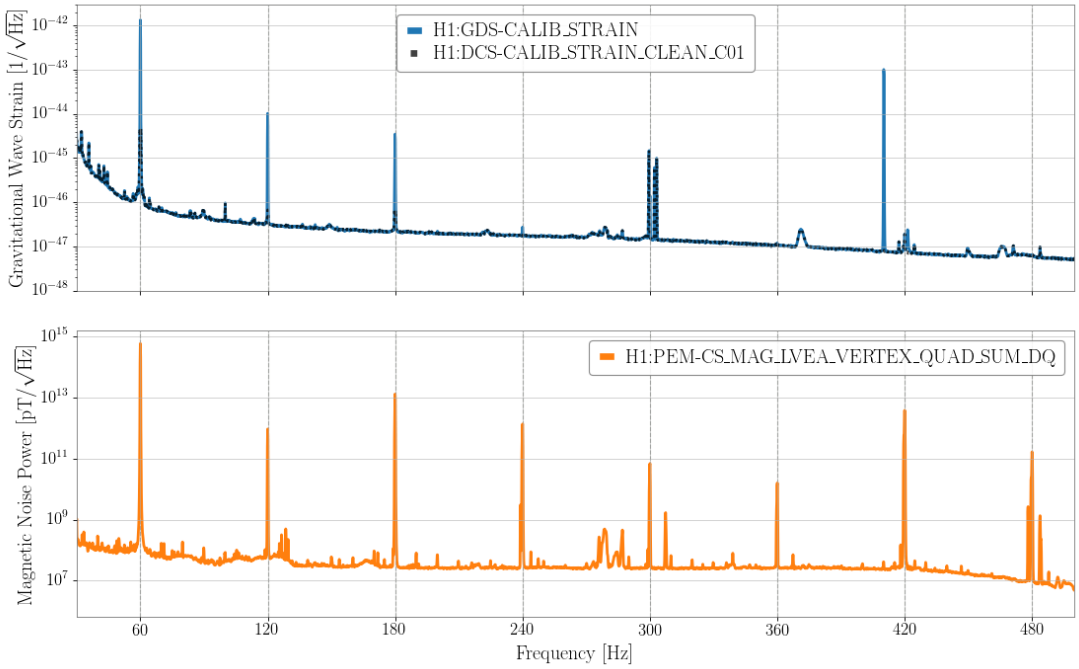

Figure 2.7: *Top:* Strain at LIGO Hanford during O3 showing the fundamental 60 Hz frequency and higher harmonics from the power lines in the DARM global diagnostic system (GDS) calibration strain channel `H1:GDS-CALIB_STRAIN`. The curve in black is the GW strain readout (blue) after filtering the electrical noise. *Bottom:* Physical environment monitor (PEM) magnetometer channel (`H1:PEM-CS_MAG_LVEA_VERTEX_X_DQ`) located at the laser and vacuum equipment area (LVEA) in the corner station (CS) of the interferometer. Since the magnetometers are able to observe the electrical noise so well, they can be used to filter the noise from the strain readout.

lines will likely always persist for ground-based detectors, the resonances from internal electronics have been almost entirely removed as of aLIGO's third observing run.

2.4 Beam Alignment and Control

The alignment of the test masses relative to each other and to the beam splitter is critical in order for destructive interference at the output port of the machine. Due to the high laser power and the fluctuation in the number of photons arriving on the

mirror surfaces, there is appreciable radiation pressure and torque on the mirror. The torque is enough to cause the mirrors to lose orthogonality with beam. This can lead to excitation of modes in the arm cavities where the mirrors twist maintaining parallelism ("soft" modes) or 180° out of phase ("hard" modes) which are unstable. Ambient vibrations can also create motion in the mirrors that couple into the length degrees of freedom. As a consequence of these motions, an alignment sensing and control (ASC) system has been put into place to drive the mirrors back into alignment. While this feedback-control loop is linear, the misalignment of the beam spot can be dynamic or static, and therefore can produce a non-stationary or nonlinear coupling to DARM. Because of this, regression where the ASC noise dominates, $\lesssim 25$ Hz, is very difficult and will be explored in detail in chapter 6.

2.4.1 Beam Jitter

In addition to the beam misalignment from the radiation pressure and subsequent ASC noise, the beam spot can become misaligned before entering the beam tube. Shaking of the table which carries the beam, or of the optics which precede the beamsplitter leads to "jittering" of the beam spot. This then couples to the longitudinal mode of the arm cavity and leads to a phase shift of the cavity light at the output port.

Beam jitter was not a particularly large problem during O3, however it was the dominant source of noise during O1 and O2 in the frequency band from $\sim 100-1000$ Hz. It was eventually determined that a water cooling pump for the high-power voltage controlled oscillator at the beam table was vibrating and causing the instability of the input beam. This cooling system was changed before the start of the third observing run alleviating the problem as can be seen in figure 2.8.

2.4.2 Light Scattering

While much work has been done to characterize the noise at and around the interferometers, figure 2.9 demonstrates that there remains unexplained power buildup particularly at low frequencies. Current efforts suggest that some of this residual power is accounted for by light that scatters out of the beam, picks up extra phase, and then recombines with the beam at or before the output port [40]. The scattering noise contributes through

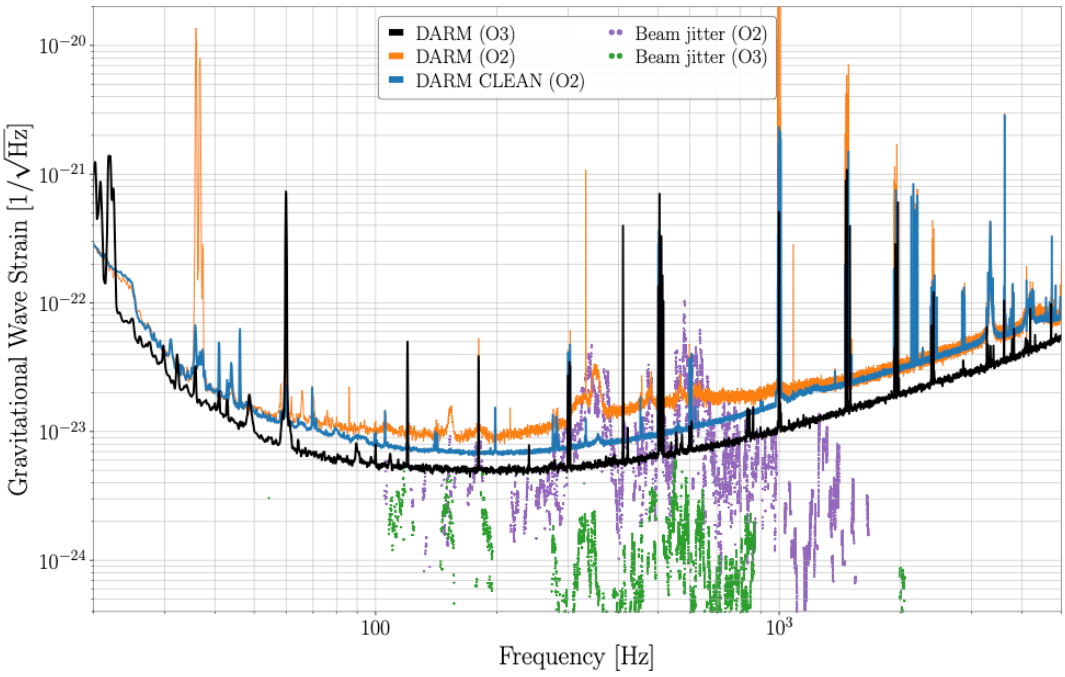

Figure 2.8: Contributions of beam jitter noise to DARM during O2 and O3. The broadband peaks in the O2 spectrum at ∼300 Hz and ∼600 Hz from the jitter noise are evident. Fortunately, this noise was linearly coupled to DARM and an optimal, linear least-mean-squares filter was used to remove the noise and produce the "DARM Clean (O2)" spectrum. The O3 jitter noise is a much more insignificant contribution to the DARM strain due to the removal of the cooling pumps causing the beam misalignment. This image was reproduced with the permission of Georgia Mendell.

"slow" and "fast" channels that span $\sim 10 - 120$ Hz. Efforts are being taken [41] to mitigate the effects of scattering and should be available as soon as O4. At the present time, it is unknown if there are other contributions to the observed broadband noise.

2.5 Auxiliary Channel Identification

The production of the noise budget shown in figure 2.2 considers only persistent noise sources. Transient noise ($\lesssim$ 1 day) will not contribute enough power to change the overall detector sensitivity after integrating over the duration of the observation period.

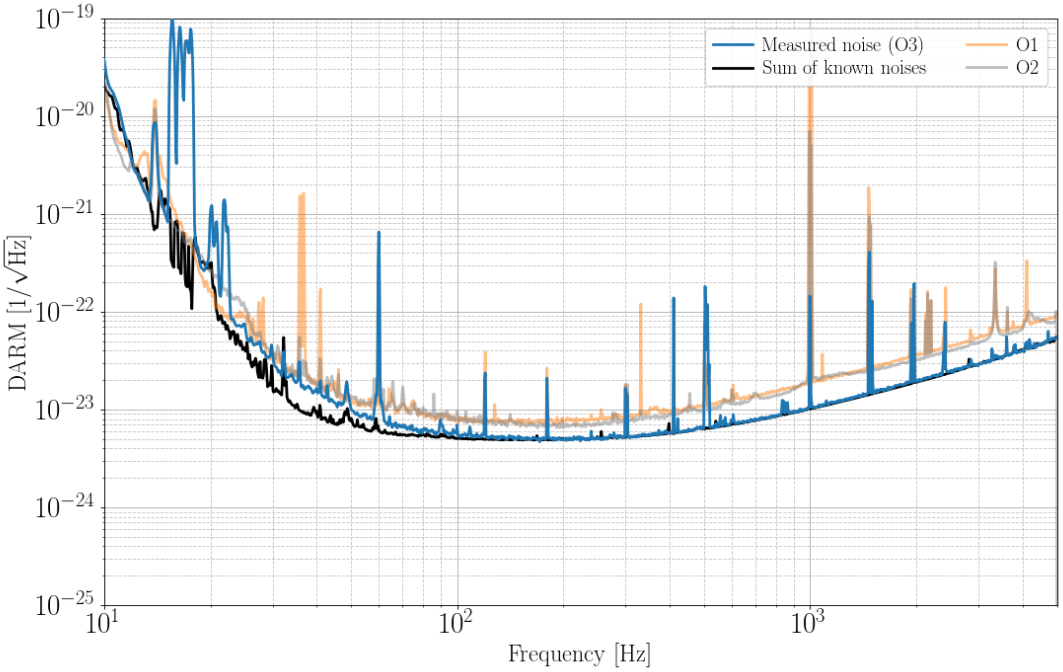

Figure 2.9: Comparison of the aLIGO sensitivities and the sum of known noises at the beginning of O3. There is a broadband region from ∼20−90 Hz which is less sensitive than expected based on the total contribution of known noise sources. Recent work suggests that this power can be largely accounted for by the scattering of light in the beam tubes [41, 40]. This image was reproduced with the permission of Georgia Mendell.

On shorter time scales however, these transients are significant enough to add narrow or broad artifacts to the power spectrum, temporarily decrease the sensitivity by orders of magnitude or to destabilize the detector enough to cause a "lock loss" period wherein the detector is not operational and must be reset. These transients often come in the form of environmental events such as high wind, thunderstorms and earthquakes although there are many more less obvious short-term sources such as sporadic instrumental noise (glitches), planes flying overhead, air conditioners or refrigerators on-site turning on, cameras clicking and phones ringing.

On time scales of $\mathcal{O}$(hours), the noise at LIGO is therefore non-stationary and non-Gaussian. The effects of the transient noise can be seen by examining a plot of the

BNS inspiral range discussed in section 2.1.2 over a few hours as seen in figure 2.10. While some noise events are severe enough that the source is immediately evident, most transient disturbances are much less obvious in appearance. This then leaves us with the task of deciphering the origin of the noise. Nominally, evidence of the noise would be caught by one of the many physical environmental monitors (PEMs) and, because the noise must have coupled into DARM, this data stream would have a significant correlation statistic with DARM. There are many methods and metrics to categorize channels with the noise origins (a variable selection and regularization method known as lasso regression [42], for example) however for the remainder of this section we shall discuss a brute-force parallelized cross-coherence software package known as the Stochastic Transient Analysis Multi-detector Pipeline for Physical Environmental Monitors, or `STAMP-PEM`.

2.5.1 Cross Coherence Measurements and `STAMP-PEM`

When an unknown noise source couples into DARM, it can be difficult to track down the cause. Although it may be possible to regress the noise in a post-processing routine, the ideal scenario would be to be able to determine the physical interaction at the detector and address the hardware setup at that location. There are many software tools which seek to identify the cause of unknown noises. Each is tailored to look at particular frequency regimes, resolutions and durations. A tool that is well-suited to quickly identify the source of spectral artifacts down to a resolution of $\mathcal{O}(10$ mHz) or greater and lasting anywhere from a few seconds to days or weeks is the stochastic transient analysis multi-detector pipeline for physical environmental monitors (`STAMP-PEM`).

`STAMP-PEM` is a software package built in python [43] that calculates coherence spectra of environmental channels coupling into DARM and automatically generates webpages of the results every two hours. The code interfaces with a high-throughput scientific computing cluster that allows calculations to be distributed across hundreds of computers at once and the results are recombined at the end. This increase in the effective speed of the code base makes is possible to use the code as a real-time monitor of the detector state and transient noise artifacts.

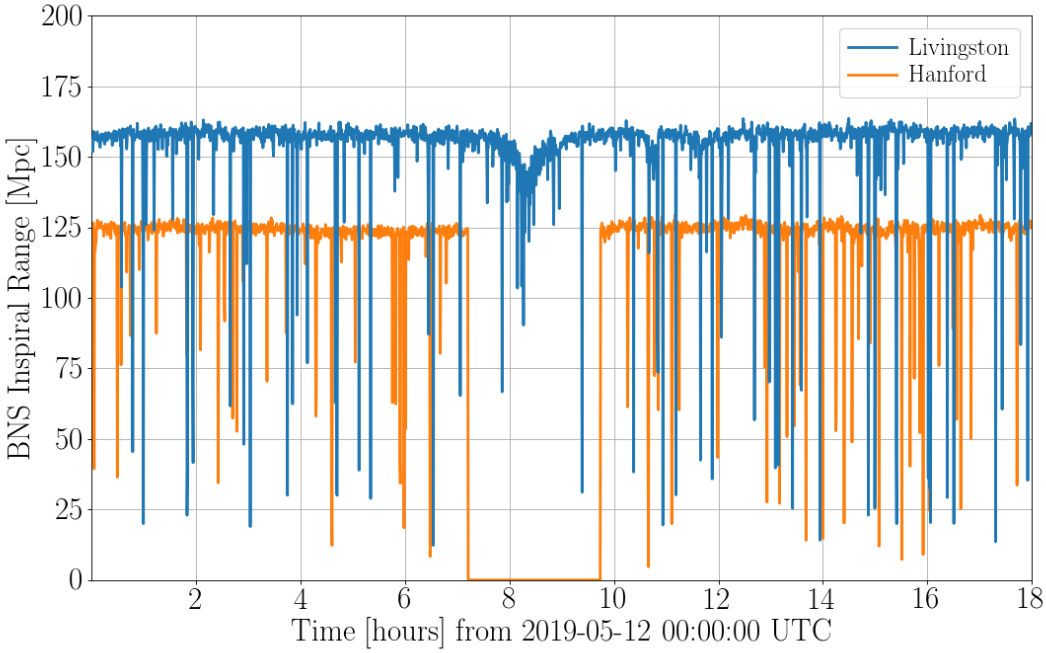

Figure 2.10: Non-stationarity of the observable range of the detectors due to noise fluctuations. While many short duration sensitivity losses are visible, a thunderstorm passing over the Livingston, LA interferometer caused a large dip in the range around 08:00 AM. Meanwhile at the Hanford, WA site, unrelated seismic noise excited the suspension wires of the test masses enough to break the lock and take the system offline.

Base Objective — The fundamental role of `STAMP-PEM` is to calculate the coherence spectrum of auxiliary environmental sensor data channels (abbreviated here to "channels") with the differential arm readout measurement at the output port of the detector. This quantitative relationship between the environmental channels and DARM provides insights about the detector status, data quality and information about how the hardware couples to the noise providing a methodological path forward for eventual upgrades. This cross-coherence is determined by

$$\Gamma(f) = \frac{\left|\mathrm{CSD}_{hw}(f)\right|^2}{P_1(h)P_2(w)} \tag{2.14}$$

where $\mathrm{CSD}_{hw}(f)$ represents the averaged cross power spectral density of the output strain channel $h(t)$ with the PEM witness channel, and $P_h(f)$ $\left(P_w(f)\right)$ is the power spectral density of DARM (witness channel). Though this is a straightforward calculation, there are $\mathcal{O}(1000)$ channels that must be run. Furthermore, the data is expected to be non-stationary which means that the coherence spectra cannot necessarily be averaged over long durations at the risk of "washing out" the transient noise. Once the coherence spectra are calculated, the results need to presented in a concise, easily referenceable way to provide at-a-glance monitoring.

Usage — The code is run during LIGO's observing runs automatically via a command-line interface. By default, coherence results are calculated at 0.1 Hz resolution and averaged over 30 minute increments. The complex CSDs, as well as the PSDs are saved in as HDF5 format. The benefit to this is any collaboration member can generate coherence data for particular channel combinations at will without needing to collect the time series and generate the coherence spectra through a companion `selective-query` tool included in the software package. Over the course of a day, the coherence spectra are averaged, saved and plotted.

Monitoring — Once at least two hours of coherence data has been collected, the data are categorized by auxiliary monitoring type (i.e., seismometers, accelerometers, hydraulic pre-isolator actuators, etc.), plotted and an interactive webpage is automatically built and uploaded to the LIGO servers. To organize the coherence spectra of any given subsystem "coherence matrices" are constructed which stack the plots of $\Gamma(f)$ for a given subsystem's channels up to ten at a time (figure 2.11). The landing page of the website consists of combined coherence matrix of every channel and subsystem as well as the detector's "locked" observing periods.

In addition to the combined and individualized plots, interactive subsystem plots built on the D3js java script library [44] are included. These interactive plots allow the user to sort an entire subsystem by channel name, frequency, coherence, or combinations thereof. Hovering the cursor over pixels gives the channel name, frequency and coherence with DARM. Because persistent noise can mask transient noise, static and interactive residual plots are created by subtracting the coherence spectra of a "typical" time

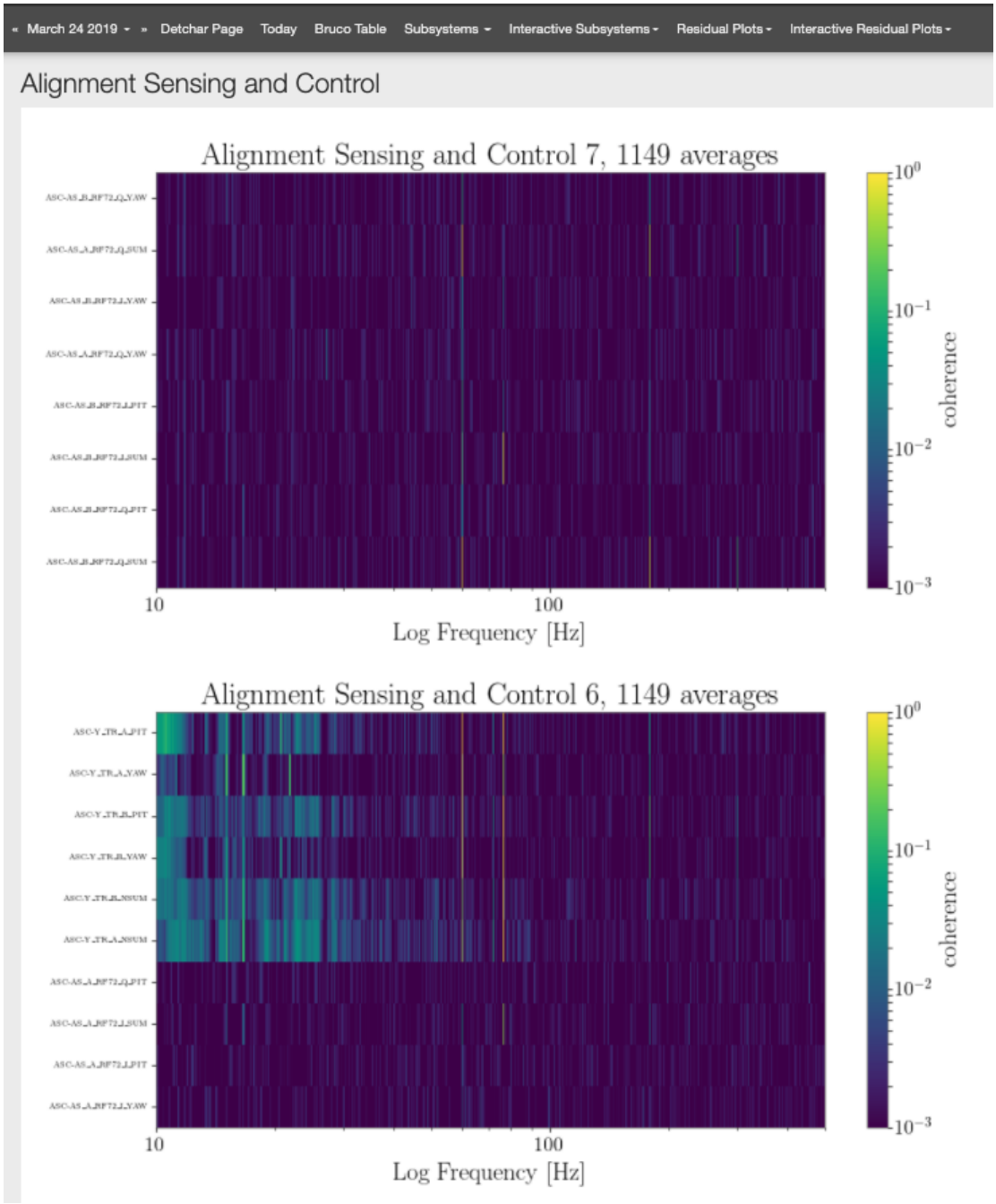

Figure 2.11: Partial output of a `STAMP-PEM` Alignment Sensing and Control subsystem page generated automatically when running the full code pipeline. Each plot is a stacked "coherence matrix" of $\Gamma(f)$ for up to ten channels per matrix. If more than ten channels exist for that subsystem, multiple coherence matrices are generated. The toolbar allows one to change the date, inspect sort-able full coherence subsystem matrices, residual plots or BruCo tables. By grouping the plots by subsystems, noise transients are much easier to diagnose and locate.

segment from the current time. This effectively removes low frequency ASC noise, 60 Hz electrical lines harmonics, violin modes of the suspension system and other unrelenting noise sources from the plots.

The result of the system-wide coherence calculations are also made available in a brute force coherence (BruCo) table, a snippet of one is shown in figure 2.12. By default, this table shows the top 20 most coherent channels in each 0.5 Hz frequency bin across three decades in frequency. The frequency resolution is adjustable down to 1 mHz, the visible frequency band can be as large or small as desired, the channel list to be analyzed may be manually chosen or inferred through regular expressions, and the number of most coherent channels to be shown may be specified. BruCo tables are extremely useful and widely used by the LIGO collaboration to diagnose the likely causes for transient lines.

Examples —

Airplane Fly Over: During aLIGO's second observing run, on March 2, 2017, spectrograms and coherence spectra from a number of subsystems including seismometers, accelerometers and microphones, showed a downward sweep in frequency from $\sim$115$-$70 Hz (figure 2.13). By cross-referencing with a plane monitoring system, it was verified that the microphone spectrograms were showing the Doppler shifted noise of a small aircraft flying over the detector airspace.

1 Hz Comb: The second observing run of aLIGO was plagued with a series of narrow spectral lines spaced at regular 1 Hz intervals across much of the frequency band from $\sim$20$-$90 Hz. The interactive coherence matrices of `STAMP-PEM` proved to be useful for monitoring the strength and changing coherence of each channel with the lines. The magnetometers showed a strong coherence indicating that the 1 Hz noise was electrical in origin as was indeed the case. Results of the PEM monitoring in the interactive plots can be seen in figure 2.14. Data collection through this pipeline was used to perform the analysis for the SGWB discussed in chapter 3.

« January 22 2017 » Today Detchar Page Subsystems Interactive Subsystems Residual Plots Interactive Residual Plots

Frequency	1 most coherent	2 most coherent	3 most coherent	4 most coherent
10.0	H1:OMC-ASC_Y2_I_OUT_DQ: 0.99893	H1:LSC-ASAIR_B_RF90_I_ERR_DQ: 0.94696	H1:LSC-DARM_OUT_DQ: 0.94524	H1:LSC-POP_A_RF9_Q_ERR_DQ: 0.91389
10.5	H1:OMC-ASC_Y2_I_OUT_DQ: 0.33465	H1:ASC-CHARD_P_OUT_DQ: 0.0529	H1:ASC-DHARD_P_OUT_DQ: 0.04796	H1:LSC-DARM_OUT_DQ: 0.04511
11.0	H1:OMC-ASC_Y2_I_OUT_DQ: 0.472	H1:ASC-DHARD_P_OUT_DQ: 0.05168	H1:ASC-CSOFT_Y_OUT_DQ: 0.04424	H1:ASC-CHARD_P_OUT_DQ: 0.04138
11.5	H1:OMC-ASC_Y2_I_OUT_DQ: 0.32572	H1:ASC-DHARD_P_OUT_DQ: 0.05673	H1:SUS-ITMY_L1_NOISEMON_UL_OUT_DQ: 0.03315	H1:ASC-CSOFT_Y_OUT_DQ: 0.03262
12.0	H1:OMC-ASC_Y2_I_OUT_DQ: 0.2964	H1:ASC-DHARD_P_OUT_DQ: 0.06462	H1:SUS-ITMY_L2_NOISEMON_LR_OUT_DQ: 0.03622	H1:SUS-ITMX_L2_NOISEMON_UR_OUT_DQ: 0.03516
12.5	H1:OMC-ASC_Y2_I_OUT_DQ: 0.33032	H1:ASC-DHARD_P_OUT_DQ: 0.07298	H1:SUS-ITMY_L2_NOISEMON_LR_OUT_DQ: 0.04785	H1:SUS-ITMX_L2_NOISEMON_UR_OUT_DQ: 0.04752
13.0	H1:OMC-ASC_Y2_I_OUT_DQ: 0.23283	H1:ASC-DHARD_P_OUT_DQ: 0.0935	H1:SUS-ITMY_L2_NOISEMON_LR_OUT_DQ: 0.05652	H1:SUS-ITMX_L2_NOISEMON_UR_OUT_DQ: 0.05617
13.5	H1:OMC-ASC_Y2_I_OUT_DQ: 0.34699	H1:SUS-SRM_M2_NOISEMON_LR_OUT_DQ: 0.19523	H1:SUS-SR3_M3_NOISEMON_UR_OUT_DQ: 0.17345	H1:LSC-DARM_OUT_DQ: 0.12677
14.0	H1:OMC-ASC_Y2_I_OUT_DQ: 0.89153	H1:SUS-ITMX_L1_NOISEMON_LR_OUT_DQ: 0.04848	H1:SUS-SRM_M2_NOISEMON_LR_OUT_DQ: 0.04288	H1:LSC-SRCL_IN1_DQ: 0.04278
14.5	H1:OMC-ASC_Y2_I_OUT_DQ: 0.31398	H1:SUS-SRM_M2_NOISEMON_LR_OUT_DQ: 0.27874	H1:SUS-SR3_M3_NOISEMON_UR_OUT_DQ: 0.26081	H1:LSC-SRCL_IN1_DQ: 0.12443

Figure 2.12: Snippet of the brute force coherence (BruCo) table output by `STAMP-PEM`. The table is arranged by ascending frequency bin on the vertical scale and ranked by the most coherent channel (with DARM) at that frequency across the horizontal axis. In addition to providing the channel names, the value of the coherence at that frequency is displayed and used to shade the entry color accordingly. In this format, locating the best (most coherent) auxiliary sensors to a particular coupling is fast and allows for narrowing down where in the detector the noise enters.

2.6 Undetermined Contamination Sources

There are a number of noise sources (known and unknown), which dictate the sensitivity of the GW detectors. The sources presented in this chapter are known to be significant and couple predominantly linearly with the detector strain. While some nonlinear and non-stationary contributions are known or suspected, they are in general poorly understood. This is in part due the unknown degree of nonlinearity of the coupling and the large amount of auxiliary data to parse through. It therefore becomes a combinatorics problem to attempt to brute force a coherence between sets of channels and DARM. It is assumed and reasonable to expect that the predominant coupling of noise within the detectors is indeed linear, and that higher order couplings will be less significant. This may be true, however the disparity in the low frequency band between known and

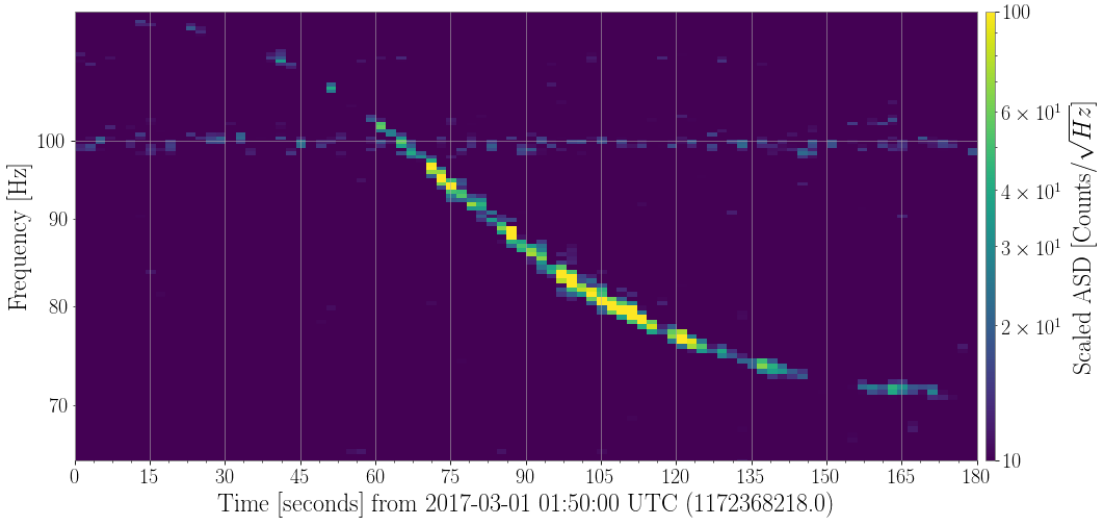

Figure 2.13: Spectrogram generated from `STAMP-PEM` of a microphone in the Hanford detector located at the corner station near the beam splitter during March of aLIGO's second observing run. The very obvious sweep in frequency was caused by a small plane flying over the detector airspace. This signal was clearly visible in the accelerometers as well. The BruCo table generated for this day was dominated in the $\sim$70$-$115 Hz band by accelerometer and microphone coherences.

measured noise as shown in figure 2.2 indicates that the spectrum is not entirely well understood. More complex couplings will be addressed in detail in chapters 5 and 6 in an attempt to solve this issue.

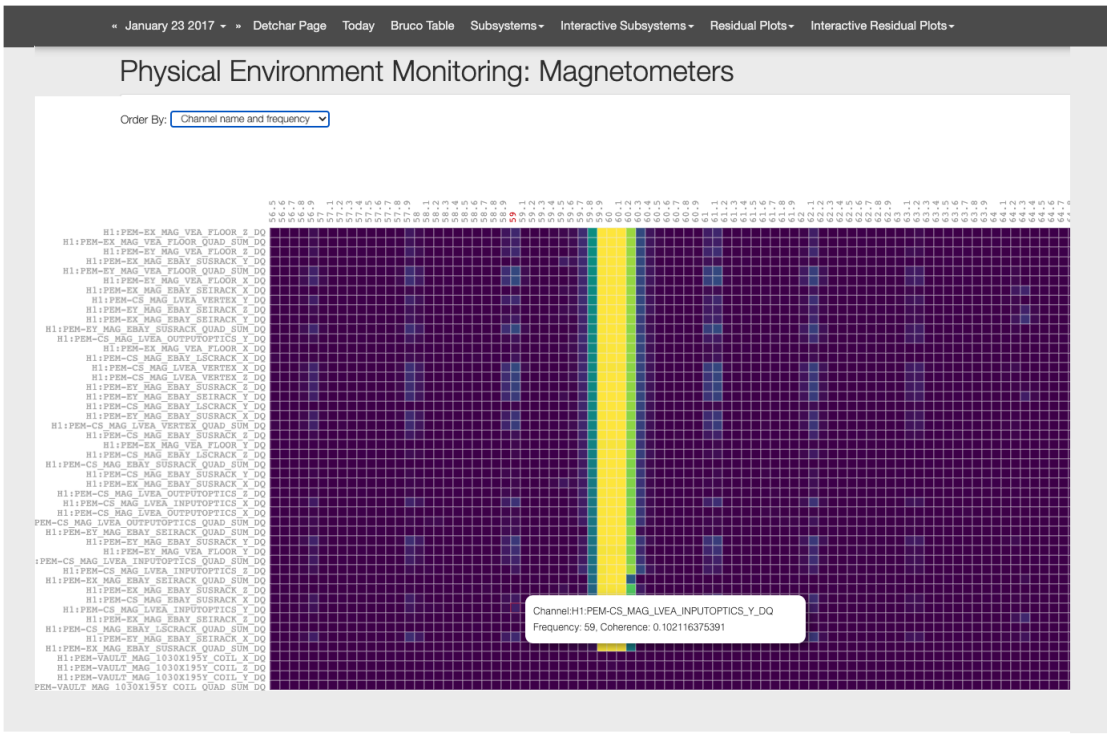

Figure 2.14: Interactive d3js plot of the magnetometers at Hanford during aLIGO's second observing run. The frequencies run along the x-axis in increments of 0.1Hz. The y-axis provides the auxiliary channel name. There is a very clear series of lines at regular 1 Hz intervals across the shown frequency band. As this is visible in nearly every magnetometer channel spanning the entire footprint of the detector site, one can quickly deduce that this noise is electrical in origin. Allowing the pointer to hover over the plot shows the information contained within the pixel including the channel name, frequency and coherence with DARM.

Chapter 3

A Cross Correlated Search for the Isotropic Stochastic Gravitational Wave Background

Gravitational waves are generated by a large variety of events including, but not limited to, compact binary coalescences (CBCs) [45, 46, 47, 48], rotating neutron stars [49, 50], supernovae [51, 52], cosmic strings [53, 54, 55] and primordial black holes [56, 57, 58]. The possibility of detection of each of these events is contingent upon the strain amplitude of the signal reaching Earth and the sensitivity of the detector. Because the signal strength falls off inversely with the distance, eventually even very high energy events are so far away that the sources themselves are no longer individually resolvable by current detectors. The superposition of these distant, unresolvable events forms a stochastic gravitational wave background (SGWB). In this chapter, we will discuss the motivations for investigating this background, its origins, as well as search methods and current upper limits from Advanced LIGO.

3.1 Motivation

The SGWB is composed in part from astrophysical sources such as BBH, BNS and BHNS inspirals. Assuming that merger rates follow the star formation rate (SFR), the

majority contribution to the astrophysical SGWB comes from sources at $z \lesssim 2$, i.e., where the SFR peaks. Because of the proximity and number of these sources, this astrophysical signal is expected to form the foreground of the total gravitational wave background (GWB). There are anticipated contributions from sources beyond $z \simeq 2$ and of cosmological origin as well, however these backgrounds are likely to be out of reach by current terrestrial detectors.

Because the coupling gravitational force to matter is ∼40 orders of magnitude weaker than even the electromagnetic force, the gravitons will have been first to decouple from interactions in the early universe - after $\sim 10^{-22}$s. We can contrast this with the earliest electromagnetic observations which can be made of the last scattering surface roughly 380,000 years after the big bang. Therefore, if a SGWB of cosmological origin can be detected, it would carry with it rich, virtually unaltered information about the conditions of the universe just fractions of a second after the coming into existence. From the information contained within this background signal, early universe models could be constrained and tested and a new window through which we observe the universe would be opened for investigation.

The astrophysical SGWB is also highly informative. The strength of the background can tell us about merger rates and star formation rates across different redshifts. The number of compact objects is in turn related to the formation of large scale structure which tells us about the expansion rate of the universe. The gravitational wave background from CBCs can consequently also constrain fundamental information about the formation and governance of the universe.

3.2 Derivation of the Isotropic Background

By definition, the events contributing to the SGWB are unresolvable and therefore the waveform of the SGWB itself in not deterministic. Instead, the SGWB is characterized only by its statistical properties. These properties are canonically contained within the normalized energy density spectrum

$$\Omega_{\mathrm{GW}}(f) = \frac{f}{\rho_c} \frac{d\rho_{\mathrm{GW}}(f)}{df} \tag{3.1}$$

where $\rho_{\mathrm{GW}}(f)$ is the energy from gravitational waves contained in a region from f to

$f + df$, and $\rho_c = 3c^2 H_0^2 / 8\pi G$ is the critical energy density required to close the universe. It is the goal of this chapter to derive the structure of $\Omega_{GW}(f)$ and to describe the isotropic cross-correlation method used to search for the signatures of this background.

It is typical to make a number of assumptions about the nature of the background. Specifically, here we assume the SGWB to be stationary, unpolarized, Gaussian and isotropic. The assumption of Gaussianity provides interpretation of the signal through the moments of the distribution and is reasonable given the large number of independent events contributing to the background energy density. Furthermore, we assume that the universe has no preferred direction and see no reason to expect one polarization over another; both conditions are supported by the homogeneity of the cosmic microwave background radiation. Lastly, we assume that the duration over which we make our measurements (< 1 yr) is small relative to the timescale on which the background varies.

Forging ahead with these assumptions, we can begin formalizing the SGWB by writing the spectral energy density of equation (3.1) in terms of the metric perturbation measured by the LIGO detectors. From equation (1.42) we know that

$$\rho_{\mathrm{GW}}(t) = T_{00}^{\mathrm{GW}}(t) = -c\frac{dE}{dAdt} = \frac{c^2}{32\pi G}\left\langle \dot{h}_{TT}^{ij}(t,\vec{\mathbf{x}})\dot{h}_{ij}^{TT}(t,\vec{\mathbf{x}})\right\rangle \tag{3.2}$$

It will prove convenient to work in the Fourier domain, and so we write the plane-wave expansion in the transverse traceless gauge as

$$h_{ij}(t,\vec{\mathbf{x}}) = \sum_{A=+,\times}\int_{-\infty}^{\infty} df \int_{\sigma} d\hat{\Omega}\, h_A(f,\hat{\Omega})\hat{e}_{ij}(\hat{\Omega})e^{-2\pi i f(t-\hat{\Omega}\cdot\mathbf{x}/c)} \tag{3.3}$$

The polarization tensors above are defined as

$$e_{ij}^{+}(\hat{\Omega}) = \hat{m}_i\hat{m}_j - \hat{n}_i\hat{n}_j \tag{3.4}$$

$$e_{ij}^{\times}(\hat{\Omega}) = \hat{m}_i\hat{n}_j + \hat{n}_i\hat{m}_j \tag{3.5}$$

where

$$\hat{\Omega} = \cos\phi \sin\theta \hat{x} + \sin\phi \sin\theta \hat{y} + \cos\theta \hat{z} \tag{3.6}$$

$$\hat{m} = \frac{d\hat{\Omega}}{d\theta} = \cos\phi \cos\theta \hat{x} + \sin\phi \cos\theta \hat{y} - \sin\theta \hat{z} \tag{3.7}$$

$$\hat{n} = \frac{d\hat{\Omega}}{d\phi}\bigg|_{\theta=\pi/2} = -\sin\phi \hat{x} + \cos\phi \hat{y} \tag{3.8}$$

To proceed, we can impose the assumptions about the background on the two-point correlation. The assumption of an unpolarized background gives

$$\left\langle h_A(f,\hat{\Omega}) h^*_{A'}(f',\hat{\Omega}') \right\rangle \propto \delta_{AA'} \tag{3.9}$$

If the background is truly isotropic, then there should be no preferred direction in space and this gives a delta function on the solid angle,

$$\left\langle h_A(f,\hat{\Omega}) h^*_{A'}(f',\hat{\Omega}') \right\rangle \propto \delta^2(\hat{\Omega},\hat{\Omega}') \tag{3.10}$$

Lastly, the stationarity of the background in frequency space may be written as

$$\left\langle h_A(f,\hat{\Omega}) h^*_{A'}(f',\hat{\Omega}') \right\rangle \propto \delta(f-f') H(f) \tag{3.11}$$

where $H(f) = H^*(f)$ and $H(f) \geq 0 \,\forall\, f$. Putting it all together we have

$$\left\langle h_A(f,\hat{\Omega}) h^*_{A'}(f',\hat{\Omega}') \right\rangle = \delta_{AA'} \delta^2(\hat{\Omega},\hat{\Omega}') \delta(f-f') H(f) \tag{3.12}$$

There is no loss of generality to assume that the SGWB has zero mean and allow the single-point correlation to vanish. If we now insert equation (3.3) into equation (3.2) and apply the conditions from equation (3.12) we find

$$
\begin{aligned}
\rho_{\rm GW}(t) &= \frac{c^2}{32\pi G}\left\langle \sum_A \sum_{A'} \int_{-\infty}^{\infty} df \int_{-\infty}^{\infty} df' \int d\hat{\Omega} \int d\hat{\Omega}'\, (-2\pi i f)(2\pi i f') h_A(f,\hat{\Omega}) \right. \\
&\qquad \left. h_{A'}(f',\hat{\Omega}') e^A_{ij}(\hat{\Omega}) e^{ij}_{A'}(\hat{\Omega}') e^{-2\pi i f(t-\hat{\Omega}\cdot\mathbf{x}/c)} e^{2\pi i f'(t-\hat{\Omega}'\cdot\mathbf{x}/c)} \right\rangle \\
&= (2\pi)^2 \frac{c^2}{32\pi G} \sum_A \sum_{A'} \int_{-\infty}^{\infty} df \int_{-\infty}^{\infty} df' \int d\hat{\Omega} \int d\hat{\Omega}'\, f f' \delta_{AA'} \delta^2(\hat{\Omega},\hat{\Omega}') \delta(f-f') \\
&\qquad H(f) e^A_{ij}(\hat{\Omega}) e^{ij}_{A'}(\hat{\Omega}') e^{-2\pi i f(t-\hat{\Omega}\cdot\mathbf{x}/c)} e^{2\pi i f'(t-\hat{\Omega}'\cdot\mathbf{x}/c)} \\
&= (2\pi)^2 \frac{c^2}{32\pi G} \sum_A \int_{-\infty}^{\infty} df \int d\hat{\Omega}\, f^2 H(f) e^A_{ij}(\hat{\Omega}) e^{ij}_A(\hat{\Omega}) \\
&= 4(2\pi)^2 \frac{c^2}{32\pi G} 2 \int_0^{\infty} df \int d\hat{\Omega}\, f^2 H(f) \\
&= 8(2\pi)^2 \frac{c^2}{32\pi G} \int_0^{\infty} df (4\pi) f^2 H(f) \\
&= \frac{4\pi^2 c^2}{G} \int_0^{\infty} H(f) f^2\, df \\
&= \int_0^{\infty} \frac{d\rho_{\rm GW}}{df} df
\end{aligned}
\tag{3.13}
$$

We may then write equation (3.12) as

$$
\left\langle h^*_A(f,\hat{\Omega}) h^*_{A'}(f',\hat{\Omega}') \right\rangle = \frac{3H_0^2}{32\pi^3} \delta_{AA'} \delta^2(\hat{\Omega},\hat{\Omega}') \delta(f-f') \frac{\Omega_{\rm GW}(f)}{f^3} \quad (f \geq 0) \tag{3.14}
$$

and since the LHS, the strain, is measurable we can in principle determine the energy spectrum of the SGWB. We must however be careful. The output from the LIGO interferometers is not $h_i(t)$, but rather $s_i(t) = h_i(t) + n_i(t)$ where $n_i(t)$ is the noise associated with the signal $s_i(t)$ in detector i. It is possible to remove this noise[1] from the analysis by implementing a cross-correlation search. Employing at least two detectors which are not co-located does allow an effective removal of noise terms since distant noise sources are, in general, uncorrelated. Use of a detector "network" introduces its own complications as the interferometers do not have a perfectly spherical antenna

[1] Only *local* noise sources are removed. Some noise sources, such as magnetic noise from resonances in the ionosphere, are not removed via cross-correlation and must be addressed separately

pattern and the imprint of the background will be measured differently at each site. For detectors which are not co-aligned or co-located we must account for the geometry and antenna pattern of each detector.

To do this, note that while the metric is a tensor, the detectors measure a scalar time-dependent strain. This scalar function is a result of the projection of the polarization tensor onto each detectors antenna pattern, $\hat{D}^{ij}$. We write

$$h(t) = h_{ij}(t, \vec{\mathbf{x}})\hat{D}^{ij} = h_{ij}(t, \vec{\mathbf{x}})\,\frac{1}{2}\left(\hat{X}^i\hat{X}^j - \hat{Y}^i\hat{Y}^j\right) \tag{3.15}$$

where X^k and Y^k give the directionality of the X and Y arms of the detector respectively. We can also define the detector response to a signal

$$F^A(\hat{\Omega}) = \hat{e}^A_{ij}(\hat{\Omega})\hat{D}^{ij} = \hat{e}^A_{ij}(\hat{\Omega})\,\frac{1}{2}\left(\hat{X}^i\hat{X}^j - \hat{Y}^i\hat{Y}^j\right) \tag{3.16}$$

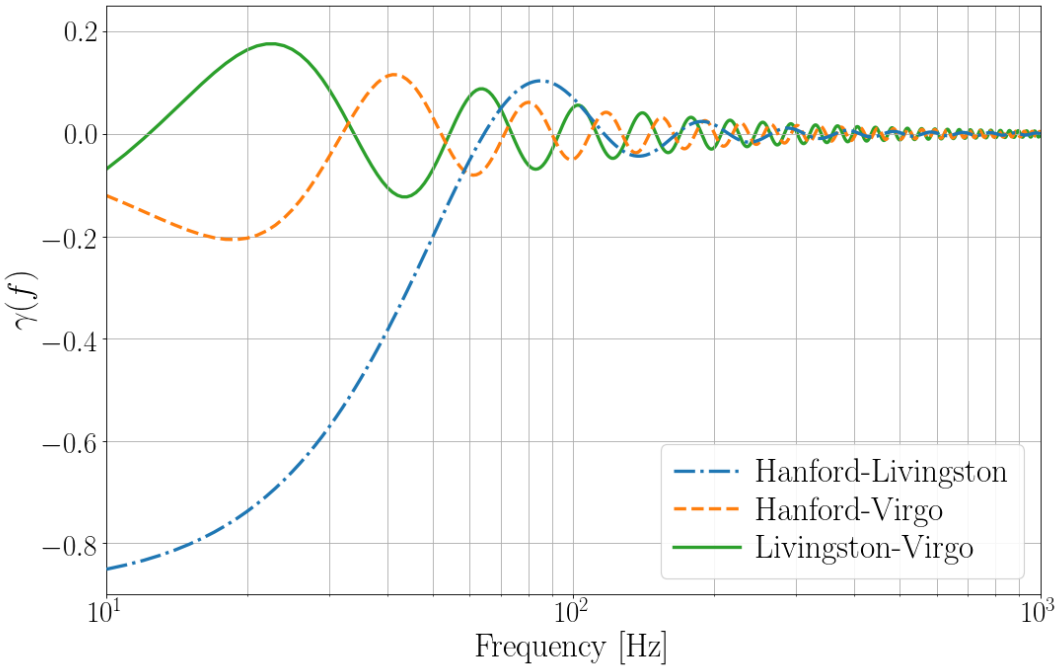

Figure 3.1: The overlap reduction function for three detector pairs.

As a final addition to this formalism, let us generalize the cross-correlation search

method by including an arbitrary filter $Q(t,t')$ which we will use to maximize the SNR of the signal. For two detectors that are neither co-aligned nor co-located and which observe for a time T, we may write

$$\begin{aligned} S &= \int_{-T/2}^{T/2} dt \int_{-T/2}^{T/2} dt' s_i(t) s_j(t') Q(t-t') \\ &= \int_{-T/2}^{T/2} dt \int_{-T/2}^{T/2} dt' \left(h_i(t) + n_i(t)\right) \left(h_j(t') + n_j(t')\right) Q(t,t') \end{aligned} \tag{3.17}$$

where $s_k(t) = h_k(t) + n_k(t)$ describes the strain of detector k as the sum of the gravitational wave signal $h_k(t)$ and noise $n_k(t)$ at that detector. Because the noise sources between the detectors are not correlated with each other or with gravitational wave signals, i.e., $\langle n_i(t) n_j(t') \rangle = 0$ and $\langle n_i(t) h_j(t') \rangle = 0$, we have

$$\mu = \langle S \rangle = \int_{-T/2}^{T/2} \langle h_i(t) h_j(t) \rangle \, Q(t) \tag{3.18}$$

The noise power is expected to be much larger that the signal power. We can therefore write the variance of the signal in the small-signal limit as

$$\begin{aligned} \sigma^2 &= \left\langle (S - \langle S \rangle)^2 \right\rangle \approx \left\langle S^2 \right\rangle \\ &= \int_{-T/2}^{T/2} dt \int_{-T/2}^{T/2} dt' \left\langle n_i(t) n_i(t') \right\rangle \left\langle n_j(t) n_j(t') \right\rangle \left| Q(t,t') \right|^2 \end{aligned} \tag{3.19}$$

and the auto correlation of the noise does not vanish. We can now calculate the mean and variance of the signal. We again use the plane wave expansion of the metric from equation (3.20), and multiply each metric tensor by the detector response shown in equations (3.15) and (3.16) to get the frequency domain strain signal. For the mean this gives

$$\begin{aligned}
\mu = \langle S \rangle &= \int_{-\infty}^{\infty} df \int_{-\infty}^{\infty} df' \, \delta_T(f - f') \tilde{Q}(f) \, \langle h_i^*(f) h_j(f) \rangle \\
&= \frac{3H_0^2 T}{16\pi^3} \sum_{A=+,\times} \int_0^{\infty} df \int d\hat{\Omega} \frac{\Omega_{\rm GW}(f)}{f^3} \tilde{Q}(f) F_i^A(\hat{\Omega}) F_j^A(\hat{\Omega}) e^{2\pi i f \hat{\Omega} \cdot \Delta \mathbf{x}/c} \\
&= \frac{3H_0^2 T}{16\pi^3} \int_0^{\infty} df \left(\sum_{A=+,\times} \int d\hat{\Omega} F_i^A(\hat{\Omega}) F_j^A(\hat{\Omega}) e^{2\pi i f \hat{\Omega} \cdot \Delta \mathbf{x}/c} \right) \frac{\Omega_{\rm GW}(f)}{f^3} \tilde{Q}(f) \\
&= \frac{3H_0^2 T}{10\pi^2} \int_0^{\infty} df \frac{\Omega_{\rm GW}(f)}{f^3} \gamma_{ij}(f) \tilde{Q}(f)
\end{aligned} \tag{3.20}$$

where we have defined the overlap reduction function (ORF), $\gamma_{ij}(f)$ as

$$\gamma_{ij}(f) = \frac{5}{8\pi} \sum_{A=+,\times} \int d\hat{\Omega} \, F_i^A(\hat{\Omega}) F_j^A(\hat{\Omega}) e^{2\pi i f \hat{\Omega} \cdot \Delta \mathbf{x}/c} \tag{3.21}$$

The ORF alone contains the integration over the solid angle and therefore encodes the geometry of each detector. The normalization out front is chosen that such that $\gamma(f) = 1$ for co-located and co-aligned detectors (i.e., when the time-of-flight between detectors vanishes: $\Delta\vec{\mathbf{x}}/c = 0$). Because this function depends upon the relative orientation of pairs of detectors, each pair will in general have its own ORF as can be seen from figure 3.1. Where the ORF is zero, the detector pair is blind to the background. This will be seen in section 3.7.

Defining the noise autopower spectrum to be

$$\left\langle n_k^*(f) n_k(f') \right\rangle = \frac{1}{2} \delta(f - f') P_k(|f|) \tag{3.22}$$

then the variance becomes

$$\sigma^2 \approx \frac{T}{4} \int_{-\infty}^{\infty} P_i(|f|) P_j(|f|) \left| \tilde{Q}(f) \right|^2 \tag{3.23}$$

We now choose the optimal filter $\tilde{Q}(f)$ which maximizes the SNR

$$\text{SNR} = \frac{\mu}{\sigma} \tag{3.24}$$

This is easily done [59] and we find

$$\tilde{Q}(f) = \xi \frac{\gamma(f)\Omega_{\rm GW}(f)}{f^3 P_i(f) P_j(f)} \tag{3.25}$$

The real, normalization constant ξ can be chosen such that $\mu = \Omega_\alpha$. The upshot of this is that the filter functions $\tilde{Q}_\alpha(f)$ are then completely determined by the detector noise spectra, the ORF, and choice of α.

Since $\tilde{Q}(f) \propto \Omega_{GW}(f)$, we must choose the filter in accordance with the background we a searching for. In other words, we either need to know what the background energy spectrum looks like or we can define a filter bank for a variety of optimal filters and search through the parameter space. We generally assume that the SGWB energy density follows a power-law distribution

$$\Omega_{GW}(f) = \Omega_\alpha \left(\frac{f}{f_{ref}} \right)^\alpha \tag{3.26}$$

As will be shown in equation (3.35), $\Omega_{GW}(f) \propto f^{2/3}$ for CBCs. The case α=0 is used to represent a cosmological background as they tend to be very nearly flat in $\Omega_{GW}(f)$ within the LIGO band ($\sim$10$-$1000Hz) [16, 19, 21]. We also employ α=3 for unmodeled events since this model is flat in GW strain and therefore weights each frequency bin equally.

Let us assemble and examine the SNR,

$$\text{SNR} = \frac{\mu}{\sigma} \approx \frac{3H_0^2 T}{10\pi^2} \left[T \int_{-\infty}^{\infty} \frac{\gamma^2(f)\Omega_{\rm GW}^2(f)}{f^6 P_i(f) P_j(f)} \right]^{1/2} \tag{3.27}$$

There are several important assessments to be made from this expression. First, the SNR is proportional to the square root of the observation time. This implies that if we observe for long enough, we will *eventually* measure a statistically significant background. Second, the model we choose for $\Omega_{GW}(f)$ is important. In other words, the overall SNR is sensitive to the power law spectral index chosen. Finally, since $SNR \propto \Omega_{GW}(f)/f^3 \propto f^\alpha/f^3$, this tells us that for a search where $\alpha < 3$, we are more sensitive to the background at lower frequencies. In the case of the gravitational wave background from CBCs, the integrand of equation (3.27) goes like $\Omega_{(2/3)}\, f^{-14/3}$ and consequently the sensitivity of the search will be dominated by low frequencies.

Finally, we can define algorithmically convenient bin-by-bin cross-correlation statistic $\hat{S}(f)$ and variance estimators $\hat{\sigma}^2(f)$

$$\hat{S}_{ij}(f) = \left(\frac{20\pi^2 f^3}{3H_0^2 T}\right) \frac{\mathrm{Re}[\tilde{s}_i^*(f)\tilde{s}_j(f)]}{\gamma_{ij}(f)} \tag{3.28}$$

$$\hat{\sigma}_{ij}^2(f) \approx \left(\frac{20\pi^2 f^3}{3H_0^2 T}\right)^2 \frac{P_i(f)P_j(f)}{8T\Delta f \gamma_{ij}^2(f)} \tag{3.29}$$

where i and j denote the detector (Hanford [H], Livingston [L], or Virgo [V]). The cross-correlation statistic (also known as the "detection statistic") is normalized so that $\left\langle \hat{S}_{ij}(f) \right\rangle = \Omega_{GW}(f)$. The estimator may be averaged over each frequency f_k to give the broadband detection statistic for each $i-j$ detector pair,

$$\hat{S}_{ij} = \frac{\sum_k w(f_k)\hat{S}_{ij}(f_k)/\sigma_{ij}^2(f_k)}{\sum_k w(f_k)^2/\sigma_{ij}^2(f_k)} \tag{3.30}$$

$$\hat{\sigma}_{ij}^{-2} = \sum_k w(f_k)^2/\sigma_{ij}^2(f_k) \tag{3.31}$$

where the weights are defined by $w(f) = \Omega_{GW}(f)/\Omega_{GW}(f_{ref})$. The final estimator is the weighted sum of each of the individual detector pair estimators,

$$\hat{S} = \frac{\sum_{ij} \hat{S}_{ij}/\sigma_{ij}^2}{\sum_{ij} \sigma_{ij}^{-2}} \tag{3.32}$$

$$\hat{\sigma}^{-2} = \sum_{ij} \sigma_{ij}^{-2} \tag{3.33}$$

3.3 CBC Contribution to the SGWB

The recent detections of BBH and BNS mergers[2] by LIGO [2, 3] allow one to estimate the rate of CBC mergers throughout the universe. The current rate estimate are $\mathcal{R}_{BNS} = 320^{+490}_{-240}$ Gpc^{-3}yr^{-1}, $\mathcal{R}_{BBH} = 23.9^{+14.9}_{-8.6}$ Gpc^{-3}yr^{-1}, and $\mathcal{R}_{NSBH} = 45^{+75}_{-33}$ Gpc^{-3}yr^{-1} [14, 60] which are higher than the initial estimates before GW detections

[2]There are two NSBH detections that will not be discussed further [60]

were made, and this in turn implies that the contribution to the SGWB from unresolvable CBC mergers will also be larger.

We may estimate the contributions to the background from CBCs in a similar fashion as in [61] and [62]. Let us assume that there is a set of averaged source parameters (spin, chirp mass, inclination etc.) given by $\vec{\theta}$, a merger rate at redshift z, $R_m(z;\vec{\theta})$ that is normalized by the local rate at $z = 0$, and that the energy spectrum at the source is $dE_{GW}(f_s;\vec{\theta})/df_s$. Including the cosmological effects, we may write the SGWB from unresolved compact binary merges as

$$\Omega_{GW}(f,\vec{\theta}) = \frac{f}{\rho_c H_0}\int_0^{z_{max}} dz\, \frac{R_m(z;\vec{\theta})}{(1+z)\sqrt{\Omega_M(1+z)^3+\Omega_\Lambda}}\frac{dE_{GW}(f_s;\vec{\theta})}{df_s} \tag{3.34}$$

At this point we can estimate the frequency dependence of the energy density of the background. Using Kepler's laws to estimate the dynamics of a CBC at low frequencies we find that

$$E = -\frac{Gm_1m_2}{2R} = -\frac{Gm_1m_2}{2}\left(\frac{\omega^2}{Gm_1}\right)^{1/3} \propto f^{2/3} \tag{3.35}$$

Therefore using equation (3.34) we arrive at

$$\Omega_{GW}(f,\vec{\theta}) \propto f^{2/3} \tag{3.36}$$

Looking back at equation (3.34), it is generally assumed that the merger rate $R_m(z;\vec{\theta})$ follows the star formation rate (SFR) $R_f(z_f;\vec{\theta})$ with some time delay distribution $p(t_d;\vec{\theta}) \propto t_d^{-1}$ as

$$R_m(z;\vec{\theta}) = \int_{t_{min}}^{t_{max}} dt_d\, R_f(z_f;\vec{\theta})p(t_d;\vec{\theta}) \tag{3.37}$$

The time delay between formation and merger for binary black holes is typically assumed to be > 50 Myr and > 20 Myr for neutron star merger delay. The upper limit of the integral, $t_{\max}$ is taken to be 13.7 Gyr. The NS masses are drawn uniformly from $(1,2)M_\odot$ whereas the primary component mass for a BBH is drawn from $p(m_1) \propto m_1^{-2.35}$ and the second is drawn uniformly such that $m_1 + m_2 < 100M_\odot$. The resulting most recent SGWB energy density spectrum from CBCs is shown in figure 3.2 and the upper limit

landscape plot across 29 decades in frequency is shown in figure 3.3.

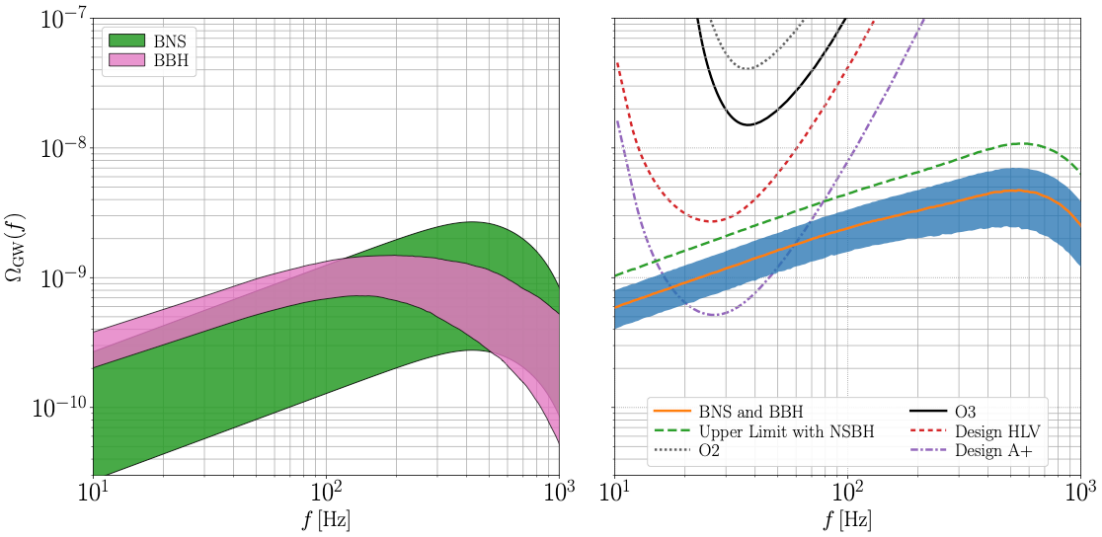

Figure 3.2: Left: 90% credible bands for the SGWB from BBHs and BNSs. While both spectra include statistical uncertainties in the merger rates, the BBH distribution also carries uncertainties in the distribution of the masses which leads to larger uncertainties at high frequencies as shown. Right: Combined BBH and BNS spectrum (orange) with the 90% credible interval given by the shaded blue region. In addition, we show the 2σ power-law integrated upper limits for O2, O3, the projected HLV 3-detector upper limits, as well as the design sensitivity for the A+ LIGO upgrades. To date there are two NSBH candidates [14, 60]. From this, a conservative NSBH rate is estimated and combined into the results to report a total upper limit to the SGWB (green dashed line). The O3 search and results will be discussed in detail in section 3.7. This figure was reproduced with permission from Tom Callister.

3.4 Detector Characterization and Data Quality

The second observing run for aLIGO extended from 16:00:00 UTC on November 30, 2016 until 22:00:00 UTC and August 25, 2017, a total period of roughly 268 days. The third observing run lasted ∼330 days in total but was subdivided into to smaller periods known as O3a and O3b which extended from 15:00:00 UTC on April 1, 2019 - 15:00:00 UTC on October 1, 2019 and from 15:00:00 UTC on November 1, 2019 - 17:00:00 UTC on March 27, 2020 respectively. Unfortunately, not all of this observation time consists

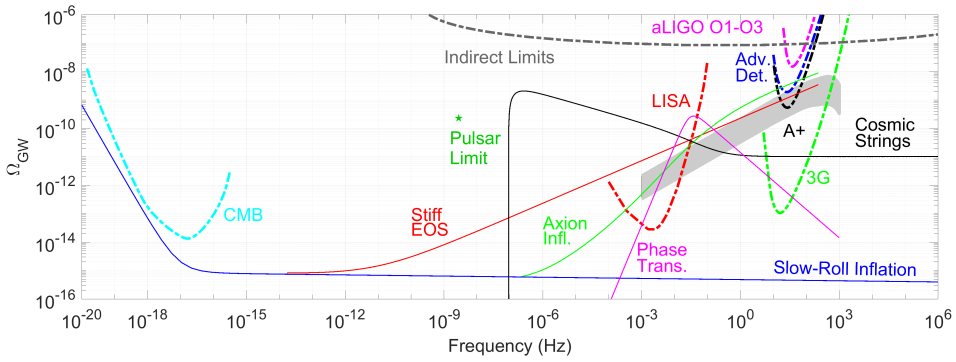

Figure 3.3: Limits on the SGWB across 29 decades in frequency [63]. The SGWB is detectable when its energy density lies above the PI curves [64]. While the inflationary background is many orders of magnitude below observation with terrestrial detectors, it can be seen that the CBC SGWB is expected to be within reach once aLIGO reaches design sensitivity. The constrains from the CMB measurements on the inflationary background however place strong restrictions at frequencies $< 30H_0$ of $\Omega_{GW}(f) < 7 \times 10^{-11}(H_0/f)^2$ [59]. See figure 3.2 for the most recent results in the boxed region. Image reproduced from [65].

of usable data. In fact, for the O3 search for the SGWB, ∼20% of the livetime was lost due to cuts in the time domain, and ∼16% of the data was removed due to frequency domain cuts (both statistics are contingent upon the detector pair). It is the goal of this section to explore the operating behavior of the detector and the environmental conditions in order to perform data quality control cuts.

3.4.1 Primary Data Cuts

LIGO utilizes a three-tier categorical ranking system to characterize the operational conditions of the interferometers. Most significant to this search are the "Category 1" vetoes. These vetoes are timestamps when the data should not be analyzed because the detector was known to be operating outside of its nominal condition or not operating at all.

In addition to losing data due to the detector status, data will also be lost for this search since any detector pair must be operational at coincident times. For O3, these combined conditions resulted in around 200 days of operational, coincident live-time for

each detector pair.

3.4.2 Non-stationarity Cuts

Until now, we have assumed that the noise within the detectors is stationary which is in general not true. Anthropogenic noise, earthquakes, high wind etc., are able to randomly and markedly move the noise floor. Because the noise floor as a whole is not stationary, we split our data into smaller segments, here we choose 192 seconds. We can then enforce stationarity within the data by removing the segments which have a standard deviation that varies by more than 20% relative to the average standard deviation of the adjacent segments,

$$0.8 < \frac{\bar{\sigma}_{\text{adjacent}}}{\sigma_{\text{center}}} < 1.2 \tag{3.38}$$

This condition is known as the $\delta\sigma$ (delta-sigma) cut and the effect can be seen in figure 3.4. In addition to reducing the non-stationarity effects of the data, this cut ensures that a transient artifact in the central segment is not overestimated due to the relatively small PSDs of the adjacent segments.

Unfortunately, the $\delta\sigma$-cut was not sufficient to address the non-stationarity of the data in O3. It was seen that exercising this condition resulted in more than half of the available data being removed. Close inspection of the segments triggering the cut revealed extremely loud, high-frequency transients with a duration $\lesssim$ 1s. These transient glitches constituted $\sim$1% of the data from the Livingston detector site and only $\sim$0.4% of the data from the Hanford site, yet they accounted for over half of the data removed by the $\delta\sigma$ cut.

To mitigate the effects, a gating method was constructed which applied an inverse Tukey window to the data surrounding the glitch [66]. The results of the gating procedure on a single glitch can be seen in figure 3.5 and figure 3.6. Using this technique, the transient glitches could be successfully removed and the segments containing the artifacts were able to be used in the analysis. We use the cross-coherence defined as a metric of the success of gating algorithm. The coherence is given by

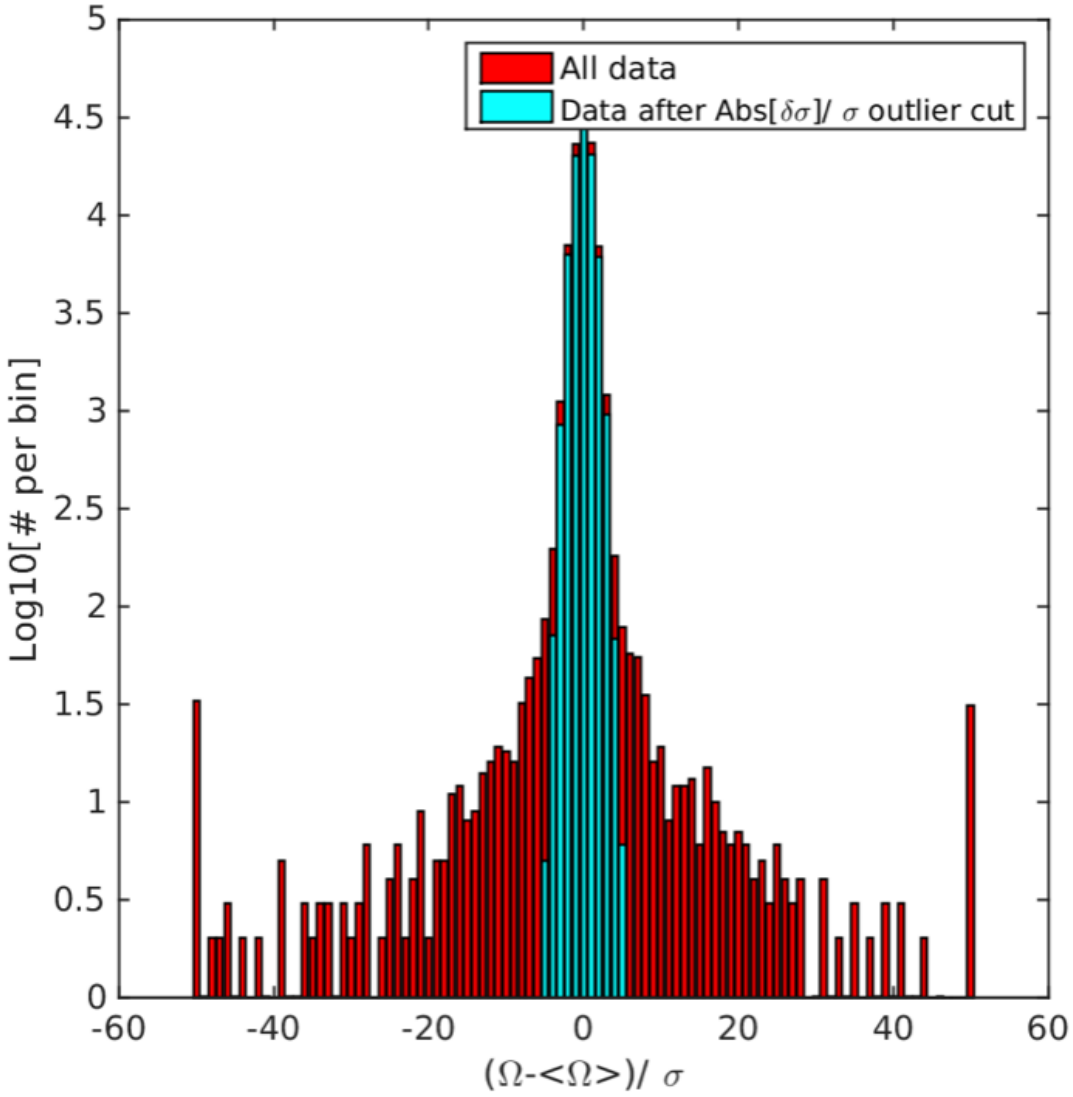

Figure 3.4: The effect of the $\Delta\sigma$ cut is to remove the segments wherein the standard deviation of the segment varies by more that 20% of the average standard deviation of the neighboring segments. This removes the tails of the distribution and the resulting standardized histogram is consistent with a Gaussian distribution.

$$\Gamma_{ij}(f) = \frac{\left\langle \left| \tilde{s}_i(f)\tilde{s}_j^*(f) \right|^2 \right\rangle}{\left\langle |\tilde{s}_i(f)|^2 \right\rangle \left\langle |\tilde{s}_j(f)|^2 \right\rangle} \tag{3.39}$$

where $\tilde{s}_x(f)$ is the Fourier transform of the time domain strain of detector x.

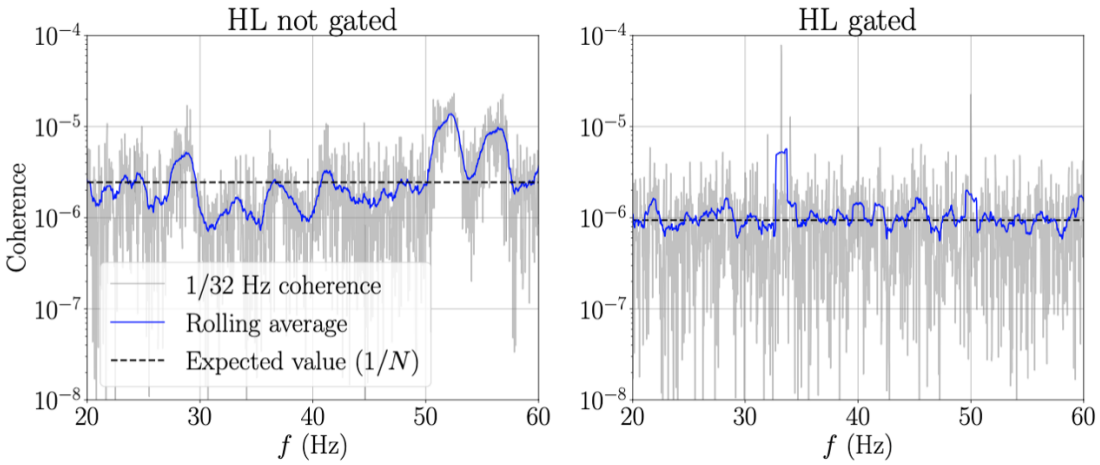

Figure 3.5: Gating effect on the Hanford-Living coherence spectra during O3. The expected coherence for random, normally distributed noise, given by the inverse of the number averages used to create the spectrum, is shown as a dashed line. After gating has been applied, the coherence spectrum is consistent with Gaussian noise. This plot was recreated with permission from Andew Matas and is used in O3 isotropic SGWB analyses paper [22].

3.5 Frequency Domain Cuts

Following the time domain vetoes and the non-stationarity cuts, we downsample the data from 16,384 Hz to 4096 Hz and apply a 16^{th}-order high pass Butterworth filter with a roll-off frequency of 11 Hz. Given the sharp rise in low frequency seismic noise, this knee-frequency removes the low frequency noise and allows one to avoid the spectral leakage from those high-power bins. Next, a Fourier transform of the data is taken on each of the segments using 50% overlapping Hann windows. The segments are then coarse grained down to 1/32 Hz resolution. While the selection of a 1/32 Hz resolution is somewhat arbitrary, it was found that this was a good compromise given the number of time-domain segments and frequency bins which needed to be removed.

Next, we assume that the coherence of the cross-correlated data follows a Gaussian distribution. Frequency bins which show excess coherence are flagged and investigated.

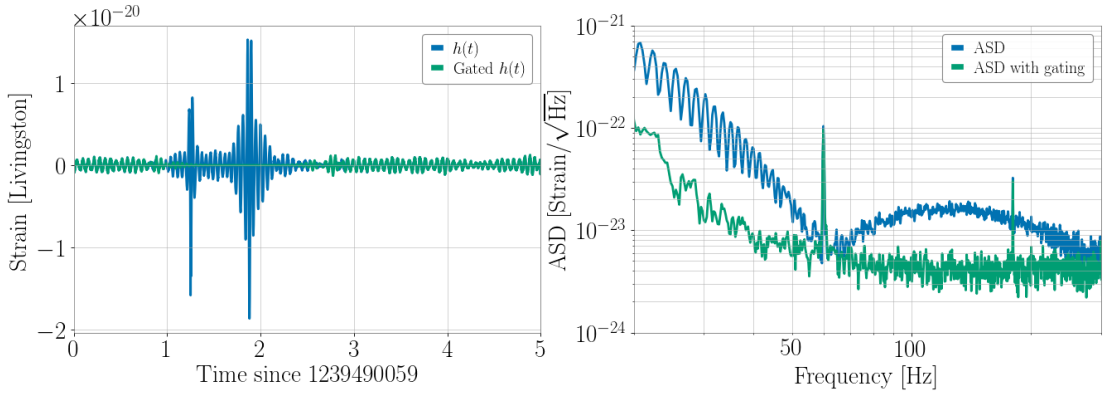

Figure 3.6: Transient high frequency glitch in the strain data at the Livingston interferometer during O3 and the removal of the glitch via inverse Tukey windowing. Although the duration of the glitch is much less than the segment duration, it is significant enough to severely bias the ASD. This plot was recreated with permission from Andrew Matas.

If this investigation reveals a source causing the excess coherence, then the frequency bin is removed from the analysis.

To perform this analysis, we first calculate the coherence as a function of frequency for the $ij-$th detector pair as given in equation (3.39). When creating a coherence from N_s segments, the probability that a coherence Γ is found is $p(\Gamma) \sim e^{-N_s\Gamma}$. With N_f frequency bins and a resolution in coherence of $\Delta\Gamma$, we may write the expected number of bins as a function of coherence as

$$N(\Gamma) = \sum_k \alpha N_f \Delta\Gamma N_s e^{-\alpha N_s \Gamma(f_k)} \tag{3.40}$$

where α is a bias factor taken to be 1 for this analysis. We set a coherence threshold where the number of bins at coherence $\Gamma(f_k)$ is equal to 1 leading to

$$\Gamma(f_k) = \frac{\log(\alpha N_f \Delta\Gamma N_s)}{\alpha N_s} \tag{3.41}$$

The results of this analysis on O3 data is shown in figure 3.7. Many of the frequencies which show an excess of coherence are a result of regularly spaced, narrow spectral artifacts known colloquially as "lines." Often, these lines are a result of the operation of electronics within the detector hardware and appear with a 0.5 Hz or 1 Hz separation

between peaks. These can clearly be seen in the raw data (figure 3.8) from $\sim$20$-$40 Hz and have largely been removed from the analysis after applying the coherence threshold cut from equation (3.41). For O3, the removal of these excessively coherent frequency bins amounted to losing 13.3%, 18.9% and 21.5% of HL, HV and LV livetime data respectively.

In addition to narrowband spectral lines from environmental coupling into the interferometer, there are similar narrow lines intentionally injected into the strain signal which need to be removed. These lines are either calibration lines as discussed in chapter 2, or pulsar injections which are used to simulate continuous wave sources. As the pulsars emit GWs, they slowly lose energy and spin down. This drift in frequency is represented in the software models and must be taken into account. To remove the frequencies which contain the pulsar injections we calculate the starting and ending frequency of the simulated pulsar and then add a buffer width of $2d$ where d is the Doppler shift and is taken to be 1×10^{-4},

$$\begin{aligned} f_{\text{start}} &= (1+d)\left[f_{\text{ref}} + \dot{f}(t_{\text{start}} - t_{ref}\right] \\ f_{\text{end}} &= (1-d)\left[f_{\text{ref}} + \dot{f}(t_{\text{end}} - t_{ref}\right] \end{aligned} \tag{3.42}$$

Since the amount of spin-down during a typical observation period is $< 1/32$ Hz, each pulsar signal is contained within just one or two frequency bins. There are generally around 20 pulsar injections during each run and therefore $\lesssim 40$ bins will be lost which is $\lesssim 0.1\%$ of all frequencies analyzed and is therefore not consequential to the overall sensitivity of the analysis.

3.6 Magnetic Noise

Although the cross-correlation between interferometers effectively removes local noise sources from the analysis, there is at least one noise source which is coherent across long distances, magnetic noise. Excitations of electromagnetic modes between the Earth's ionosphere and surface, known as Schumann resonances, create fields with very long

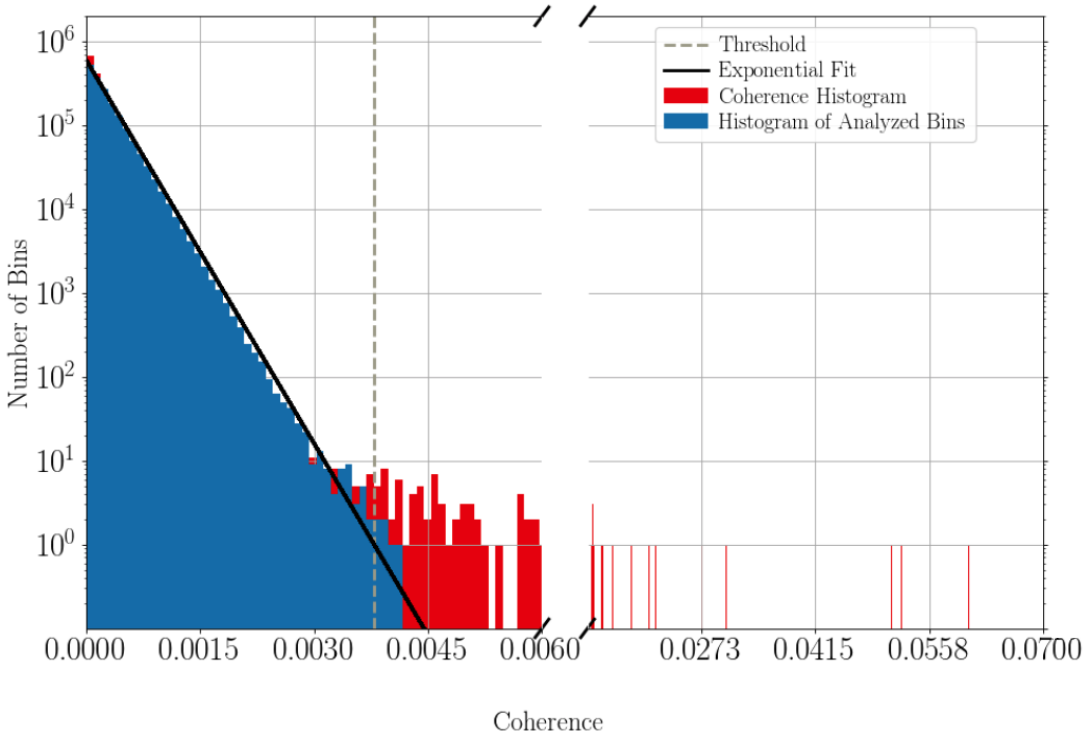

Figure 3.7: Raw coherence distribution (red) and the expected coherence assuming Gaussian noise (black line) on O3 Hanford-Livingston data with 1mHz resolution frequency bins. The blue histogram is the data left over after removing the bins with excess coherence and a known coupling.

coherence lengths and which travel at the speed of light. This is similar to the nature of the SGWB we endeavor to detect and this contribution must therefore be well understood in order that we do not conflate a magnetic correlation for a gravitational one.

Fortunately, it is possible to estimate the total magnetic noise budget. To do this, we use sensitive LEMI (Laboratory of Electromagnetic Innovations) magnetometers at each site. Using the magnetometer output data $m_k(t)$, we calculate the real part of the cross power spectrum between detectors ij, $\mathrm{Re}\,[\tilde{m}_i^*(f)\tilde{m}_j(f)]$. Next, we estimate the transfer function between the magnetic contamination and the coupling into each interferometer, $T_k(f)$, through linear interpolation. This is accomplished by performing weekly magnetic injections at each detector site and measuring the detector response to

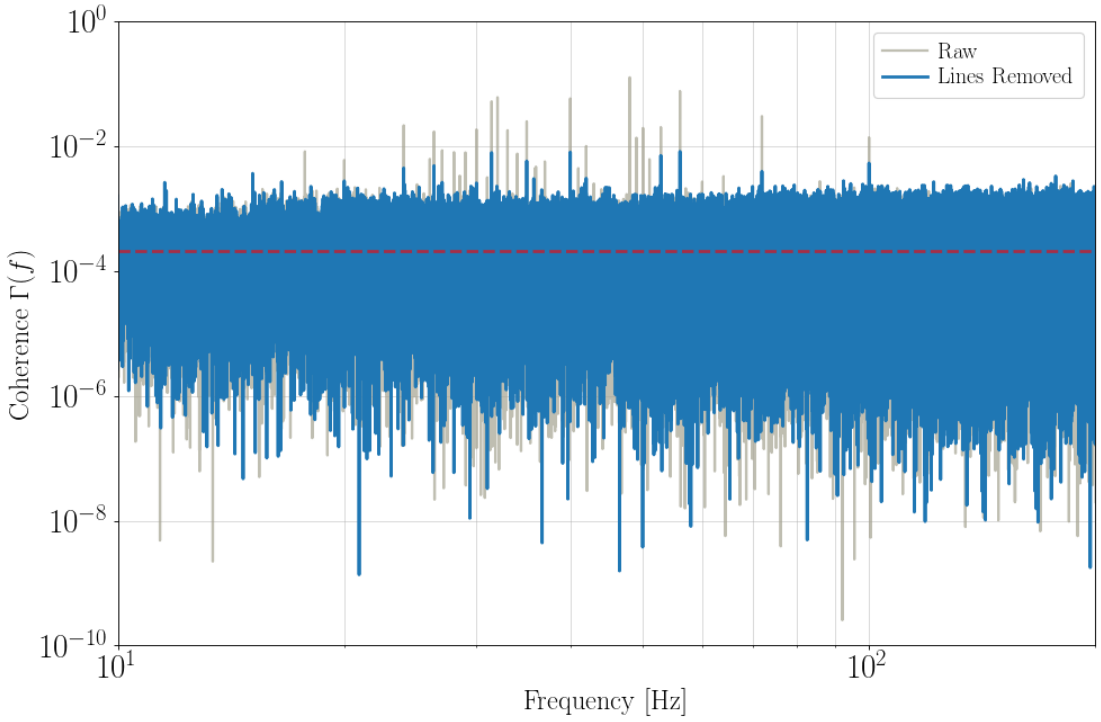

Figure 3.8: Application of the frequency domain coherence threshold cuts for the isotropic SGWB search on O3 data from the Hanford-Livingston detector pair. In the raw cross-correlated data, regularly spaced lines can be seen in the low frequency band. After inspecting these low frequency lines, many are found to be a result of known environmental couplings and are then removed from the analysis. This modified coherence spectra is shown as "Lines Removed" above. The dashed red line is the expected coherence for uncorrelated Gaussian noise given the number of segments (here $N_s = 4859$). In addition to the removal of the high coherence frequency bins, we have also removed the calibration lines, pulsar injections and known linear couplings from the data.

a known injected field [67]. Accounting for the geometry of the $i-j$ detector pair with the ORF and accounting for proper normalization, we estimate the magnetic energy density spectrum as

$$\hat{\Omega}_{ij}^{\mathrm{mag}}(f) = \frac{20\pi^2 f^3}{3H_0^2 T} \frac{|T_i(f)||T_j(f)\mathrm{Re}\left[\tilde{m}_i^*(f)\tilde{m}_j(f)\right]}{\gamma_{ij}(f)} \tag{3.43}$$

Once each detector pair's magnetic noise budgeted is computed, we combine the

baselines using a weighted statistic analogous to equation (3.31). We find that for the O3 SGWB analysis, the magnetic noise budget is still well below the current sensitivity limit (see figure 3.9) and is not considered further. This will not be the case for future observations for LIGO; after the A+ upgrades are made, the magnetic noise will be comparable to the sensitivity of the experiment to the stochastic background and must be considered more thoroughly. Current efforts are underway to perform an optimal least squares linear filtration of the magnetic noise from the cross correlated strain data [68, 69, 70].

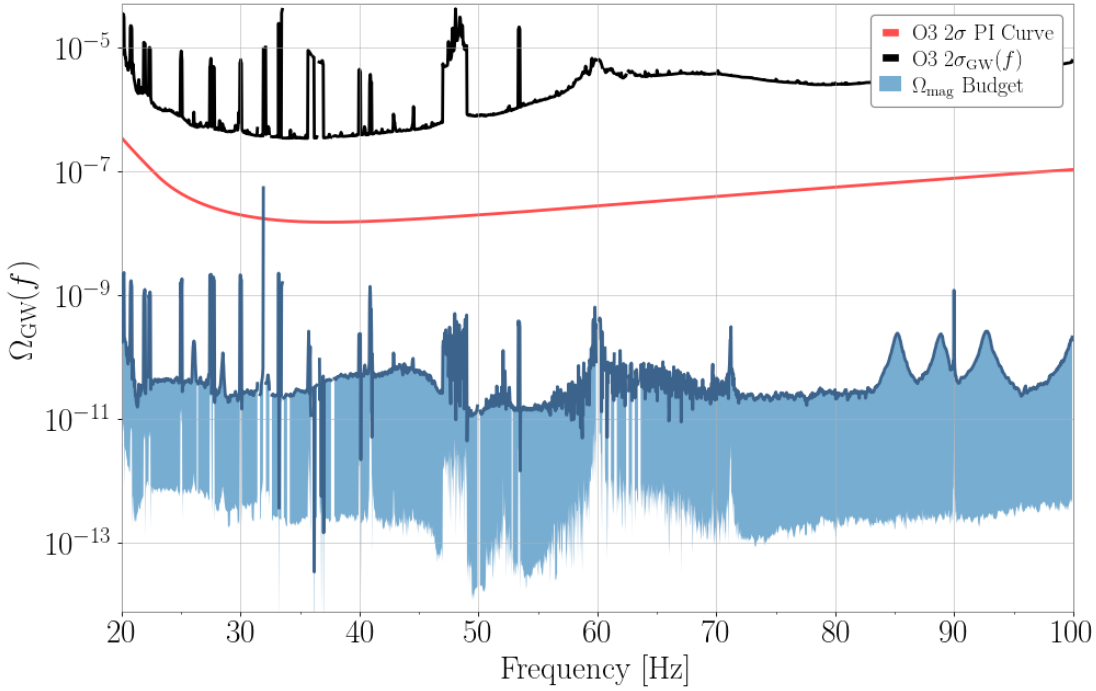

Figure 3.9: O3 correlated magnetic noise budget. The light blue band shows the weekly-averaged expected range of magnetic contamination and accounts for the uncertainly in the measurements of the coupling function. The red curve is the power-law integrated (PI) curve, which gives the sensitivity of the search to power-law backgrounds and includes the effect of integrating over frequencies. It should be compared with the overall trend of the magnetic noise budget whose upper limit is shown in dark blue. The square root of the variance, $\sqrt{\sigma^2_{GW}(f)}$, is shown as a black line, which gives the sensitivity of the search to narrowband features. This plot has been reproduced using the data from [22].

3.7 Current Upper Limits on the SGWB

In this section we present the current results for the isotropic gravitational wave background search using the combined O1+O2+O3 data. We quote results for three different astrophysical models implemented though alternate choices of spectral index in the power law model of equation (3.26):

- $\alpha = 0$. This model describes the cosmological contribution to the SGWB through slow roll inflation, cosmic strings and first order phase transitions
- $\alpha = 2/3$ corresponds to the astrophysical background comprised of the superposition of unresolved compact binary mergers
- $\alpha = 3$ is a model which is chosen since it is flat is the strain power spectrum and therefore weights each frequency bin equally. This choice is then appropriate for sources which are unmodeled, or are expected to produce GWs approximately equally across a large frequency band.

The results for the isotropic SGWB are consistent with Gaussian noise (see figure 3.10) and since no signal was detected, we place upper limits on the energy density of the background assuming the power law model and spectral indices above. The point estimate and 1σ error bars for each of the three detector pair combinations, Hanford-Livingston (HL), Hanford-Virgo (HV), and Livingston-Virgo (LV) are presented in table 3.1. We also show the frequency band which contained 99% of the sensitivity for the given model. Since SNR $\propto \Omega_{GW}(f)/f^3 \propto f^\alpha/f^3$, the smaller spectral indices weight lower frequencies more heavily and thus the majority of the search sensitivity is contained within a lower frequency band. It may be noted that the HV and LV baselines have a mean which is $\sim 2\sigma$ away from zero, however the much more sensitive HL statistic *is* consistent with a mean of 0 and we consequently do not consider this deviation as evidence of a detection.

Although we cannot claim a detection, we may determine new upper limits on the amplitude of the power law model, Ω_α, for the spectral indices $\alpha = \{0, 2/3, 3\}$. We begin by assuming a normally distributed likelihood for the cross correlations,

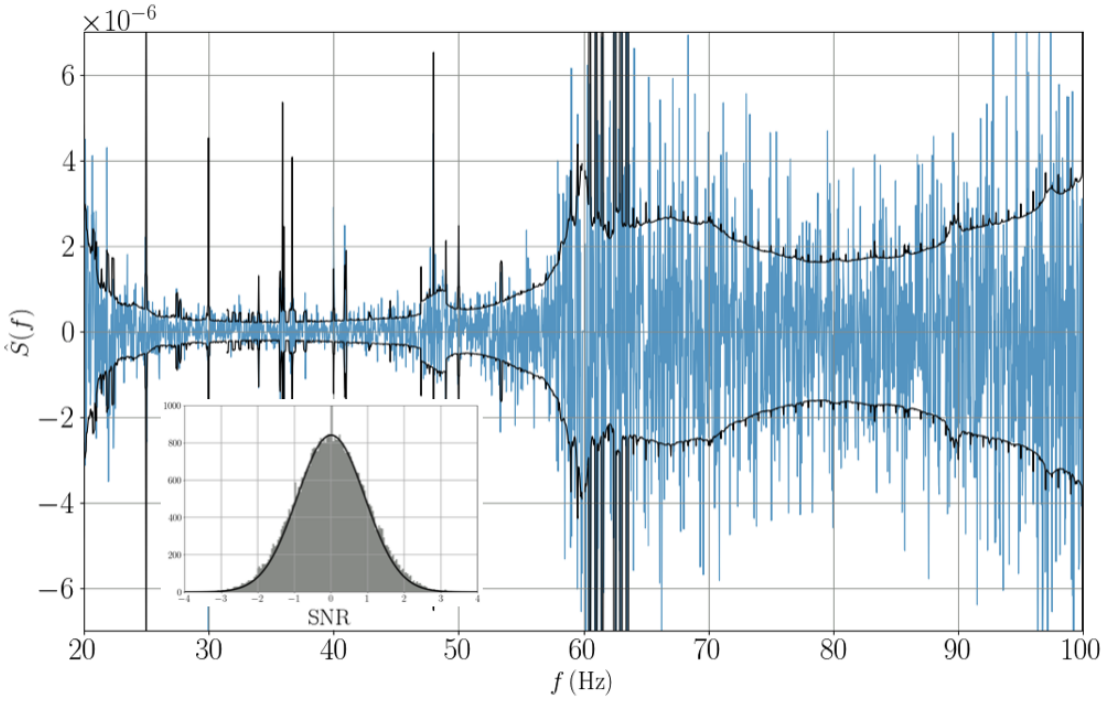

Figure 3.10: Bin-by-bin estimator for the O1+O2+O3 combined isotropic SGWB search. The black envelope is the 1σ error bar. The loss in sensitivity near 60 Hz is due to a zero in the overlap reduction function. The overlayed histogram shows the binned SNR as well as a Gaussian profile with zero mean and a standard deviation of 1 showing that the results are consistent with normally distributed noise.

$$\mathcal{L}(S|\Theta) \propto \exp\left(-\frac{1}{2}\sum_k \frac{(S(f_k) - \langle S(\Theta; f_k)\rangle)^2}{\sigma^2(f_k)}\right) \tag{3.44}$$

where $\langle S(\Theta; f_k)\rangle$ is the cross correlation one would expect at frequency f_k given the model parameters Θ. We may combine the results from the O1, O2 and O3 observing runs by using the joint likelihood given by multiplying the individual likelihoods,

$$\mathcal{L}\left(C_{O1}, C_{O2}, C_{O3}|\Theta\right) = \mathcal{L}(C_{O1}|\Theta)\mathcal{L}(C_{O2}|\Theta)\mathcal{L}(C_{O3}|\Theta). \tag{3.45}$$

We now choose two prior probability distributions on Ω_α - one which is uniform in the strength of the gravitational wave background ($p(\Omega_\alpha) \propto 1$), and one which is

Power law	$\hat{S}^{HL}/10^{-9}$	$\hat{S}^{HV}/10^{-9}$	$\hat{S}^{LV}/10^{-9}$	$f_{99\%}^{\mathrm{O1+O2+O3}}$ [Hz]	$\hat{S}^{\mathrm{O1+O2+O3}}/10^{-9}$
0	-2.1 ± 8.2	229 ± 98	-134 ± 63	75.6	1.1 ± 7.5
2/3	-3.4 ± 6.1	145 ± 60	-82 ± 40	90.6	-2.5 ± 5.6
3	-1.3 ± 0.9	9.1 ± 4.1	-4.8 ± 3.1	291.6	-0.6 ± 0.8

Table 3.1: Search results for an isotropic gravitational wave background using the optimal filter method for power law with $\alpha = \{0, 2/3, 3\}$. The point estimate and 1σ uncertainty for the cross-correlation estimate $\hat{S}_{ij}$ are shown for each $i-j$ detector pair (HL, HV, LV). We also show the frequency band from 20 Hz to $f_{99\%}$ containing 99% of the sensitivity. In the last column, we show the search results after combining the data from all three aLIGO observing runs (O1, O2, O3). While the point estimates for the HV and LV baselines are approximately 2σ away from zero, the HL configuration is consistent with zero. Since HL is the most sensitive detector pair, we do not take the deviation from zero in HV and LV to be physically significant.

α	Uniform Prior	Log-uniform Prior
0	1.7×10^{-8}	5.8×10^{-9}
2/3	1.2×10^{-8}	3.4×10^{-9}
3	1.3×10^{-9}	3.9×10^{-10}
Marg.	2.7×10^{-8}	6.6×10^{-9}

Table 3.2: Upper limits at the 95% credible level on Ω_α under the power law model for the SGWB given different spectral indices. Also shown is the marginalized limit obtained by integration over α, using a Gaussian prior with $\mathcal{N}(\mu{=}0, \sigma{=}3.5)$. The uniform upper limits are more conservative, while the log uniform priors are more sensitive to weak signals.

uniform in the log of the strength of the background ($p(\Omega_\alpha) \propto \Omega^{-1}$). Multiple priors are used as the uniform prior gives a more conservative upper limit and the log-uniform prior is more sensitive to weak signals. The joint likelihood is then marginalized over each of the priors separately to produce the combined upper limits. The results are listed in table 3.2.

These new upper limits represent an improvement from the log-uniform prior of $\sim\{6.0, 8.8, 13.1\}$ and $\sim\{3.6, 4.0, 5.9\}$ from the uniform prior for $\alpha = \{0, 2/3, 3\}$ respectively. The improvement of the isotropic SGWB search where $\alpha = 3$ is in part due to the detector upgrades (such as frequency-dependent squeezing [71]) which enhanced the sensitivity in the high frequency regime.

Chapter 4

The Cosmic Explorer Astrophysical Foreground

4.1 Introduction

One of the most exciting prospects for third generation (3G) gravitational wave instruments is the detection of a cosmological gravitational wave background (CGWB). This background may carry with it valuable information about the conditions of the early universe that constrains the prevalence and contribution of early universe GW phenomenology such as cosmic strings, primordial black holes, first order phase transitions and inflationary models. The astrophysical background of CBCs from distant galaxies would carry rich information about the rate of mergers from compact objects, their distribution, the formation of large scale structure in the universe, and much more.

In this chapter we shall consider the 3G detector "Cosmic Explorer" and investigate how we may use this new detector data stream to characterize the gravitational wave background (GWB). Given the rate of CBCs in the observable universe, it is likely that the GWB will be in part masked by the astrophysical events in the foreground. Consequently, we will address the significant and novel challenges of handling this astrophysical foreground of overlapping BNS and BBH signals. The workflow guiding this analysis is subdivided as follows:

1. Create two noise realizations consistent with the proposed CE detector sensitivities [72, 73]
2. Define CBC parameter distributions using accepted astrophysical models and current rate estimates
3. Draw CBC event parameters, create the corresponding signals and add them to the mock detectors output data streams
4. Simulate and add a GWB to each detector
5. Estimate the CBC events using the chosen parameters
6. Use the estimate of the events to create binary masks
7. Apply the event masks to the cross-correlated data in the frequency domain and attempt to recover the GWB

By including (5), we are assuming that 3G parameter estimation pipelines will be able to uniquely untangle overlapping waveforms. This is a nontrivial point. In section 4.6 we will briefly discuss one possible way around this obstacle.

4.2 Cosmic Explorer

The CE detector is a proposed 3G ground-based interferometer with a similar "L-shaped" layout like the LIGO, Virgo and Kagra detectors. The initial hardware configuration will mirror that of the updated aLIGO (A+) upgrades [72]. The primary configuration difference is that the cavity arms will be 40km long instead of the 4km currently used by aLIGO. Because the strain sensitivity in the interferometers goes like $\Delta L/L$ where L is the length of the cavity arms, we would roughly expect an increase in the sensitivity of CE to A+ of a factor of $\sim$10. This may not be entirely true however since it is not expected that all noise sources will scale trivially with the length of the cavity arms [73]. Ignoring these subtleties for now, we may plot and compare the sensitivities of current and future detectors as in figure 4.1.

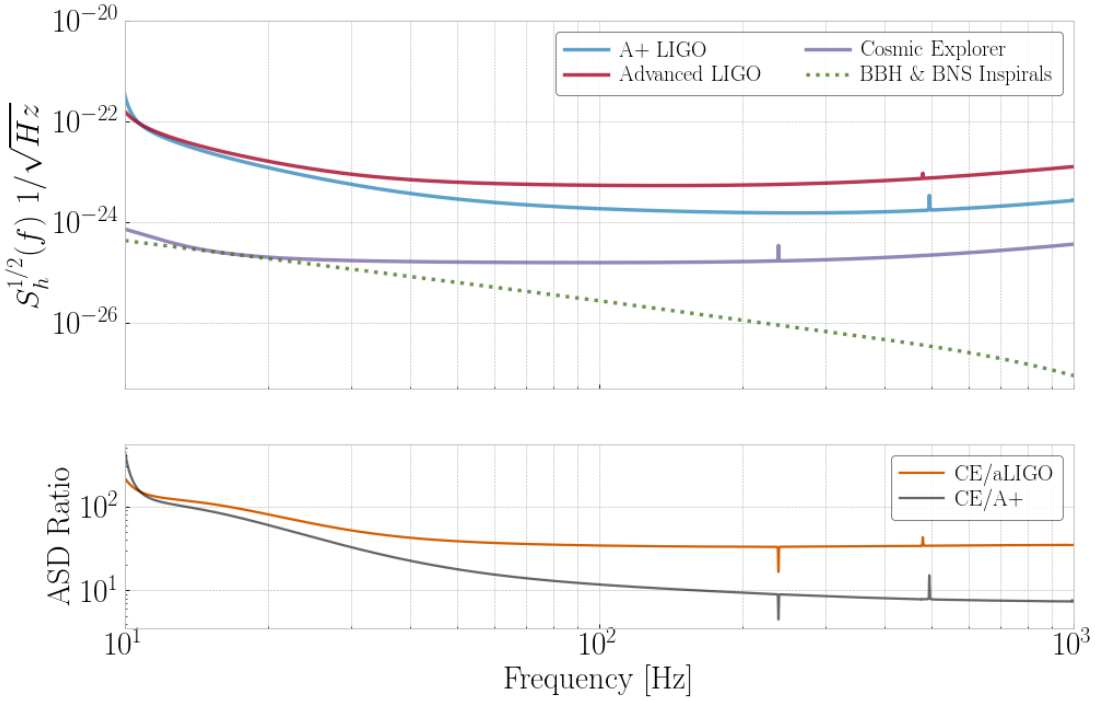

Figure 4.1: *Top panel*: Strain sensitivities for aLIGO, A+ (estimated sensitivity of aLIGO after final upgrades) and Cosmic Explorer [74, 75, 76]. Also shown is the estimated strain from BBH and BNS coalescences estimated using the rate and population data from the O3 isotropic SGWB analysis [22]. *Bottom panel:* The ratio of the ASDs of Cosmic Explorer to aLIGO and A+ are seen here. As expected, CE is approximately 10x more sensitive than A+ due to the extended cavity arms and roughly 30x more sensitive than what is currently achieved by aLIGO.

A few comments about figure 4.1 are in order. First, we see that indeed the design sensitivity of CE relative to A+ is improved by about an order of magnitude. Second, if we note that

$$\text{SNR} \propto \frac{(h_0)_{\text{min}}}{\sqrt{S_h(f)}} \tag{4.1}$$

where $(h_0)_{\text{min}}$ is the minimum amplitude that can be detected at a particular SNR, and $S_h(f)$ is the PSD of the detector's noise floor, (see [19]), then

$$\mathrm{SNR} \propto r_{\mathrm{BNS}}^2 \propto \frac{1}{\sqrt{S_h(f)}} \tag{4.2}$$

where r_{BNS} is the BNS inspiral range as described in equation (2.2). From this we can interpret the ratios in the bottom panel in two useful ways: either as the ratio of the improvement of the detector's observable distance, or as fractional increase in sensitivity of the detector. Third, we see that the combined contribution of BBH and BNS events will limit the sensitivity of CE at low frequencies. This limitation is severe and nontrivial. The majority of this chapter will detail a method which simulates the population of CBCs observed by CE and subsequently estimates and removes the resolvable events from the data thereby exposing potential other GW backgrounds.

4.3 Data Generation

As a first step, we create a noise realization simulating CE sensitivities. To do this, we multiply the CE ASD by white noise and take an inverse Fourier transform into the time domain. In a similar fashion, we may define and add a signal for the GWB. We choose a power-law form given by

$$\Omega_{GW}(f) = \Omega_\alpha \left(\frac{f}{f_{\mathrm{ref}}}\right)^\alpha \tag{4.3}$$

where we choose $f_{\mathrm{ref}} = 25$ Hz. Since we will be performing a cross-correlation search, we must create the same realization of the GWB into two interferometer simulations. To this end, we first convert the power-law above into the strain power spectrum

$$S_{GW}(f) = \left(\frac{3H_0^2}{10\pi^2}\right) \frac{\Omega_{GW}(f)}{f^3} \tag{4.4}$$

This strain is then multiplied by white noise and added into each detector using the overlap reduction function (ORF) $\gamma(f)$,

$$
\begin{aligned}
z_i(f) &= \sqrt{S_{GW}(f)} * w_n^{(i)} & z_j(f) &= \sqrt{S_{GW}(f)} * w_n^{(j)} \\
\xi_{1,ij}(f) &= \left(1 + \sqrt{1-\gamma_{ij}^2(f)}\right)^{1/2} & \xi_{2,ij}(f) &= \gamma_{ij}(f)\left(1 + \sqrt{1-\gamma_{ij}^2(f)}\right)^{-1/2} \\
h_i(f) &= \frac{\xi_{1,ij}(f) z_i(f) + \xi_{2,ij}(f) z_j(f)}{\sqrt{2}} & h_j(f) &= \frac{\xi_{2,ij}(f) z_i(f) + \xi_{1,ij}(f) z_j(f)}{\sqrt{2}}
\end{aligned} \tag{4.5}
$$

where $w_n^{(k)}$ is a random realization of white noise with as many bins as $z_k(f)$, $\gamma(f)_{ij}$ is the ORF for the ij detector pair, and $h_k(f)$ is the frequency-domain stochastic background signal added into detector k. Notice that $\xi_{1,ij}^2 + \xi_{2,ij}^2 = 1$ and from a geometric point of view these terms are effectively sine and cosine mixing angles. Here, we will consider the case of a cosmological background which is flat in $\Omega_{GW}(f)$, i.e., $\alpha = 0$.

Next, we need to estimate and add the expected contribution of CBCs. Roughly speaking, this involves picking astrophysical distributions for each set of parameters and then randomly drawing parameters from each distribution and constructing the corresponding CBC signals. The next several subsections will cover these selection processes.

4.3.1 Rates and Population

The BBH and BNS merger rates used for this analysis come from the results quoted in LIGO's First Gravitational Wave Transient Catalog [2], although more recent results can be found in the second GW catalog [14]. Specifically, we adopt the mean results $R_{\rm BNS} = 10^3\ {\rm Gpc}^{-3}\,{\rm yr}^{-1}$ and $R_{\rm BBH} = 55\ {\rm Gpc}^{-3}\,{\rm yr}^{-1}$. Therefore BBH mergers comprise just 5.5% of the expected background giving rise to the "popcorn background" shown in figure 4.2. From these rates, we can quickly estimate the number of events we should expect per second. Assuming that CE can measure out to $z{\sim}2$, then the distance out to which we observe is ∼51.7 billion light years, this gives a volume of $\mathcal{O}(10^3)$ Gpc^3. With a rate of $\sim 10^3\ {\rm Gpc}^{-3}\,{\rm yr}^{-1}$, we find

$$
N_{\rm events} {\sim} \frac{10^3\,{\rm events}}{{\rm Gpc}^3\,{\rm yr}} \times 10^3 {\rm Gpc}^3 \times \frac{1\,{\rm yr}}{10^7 {\rm s}} \sim \frac{1\,{\rm event}}{10 {\rm s}} \tag{4.6}
$$

Next, we assume that the merger rate of events, $R(z)$, follows the star formation

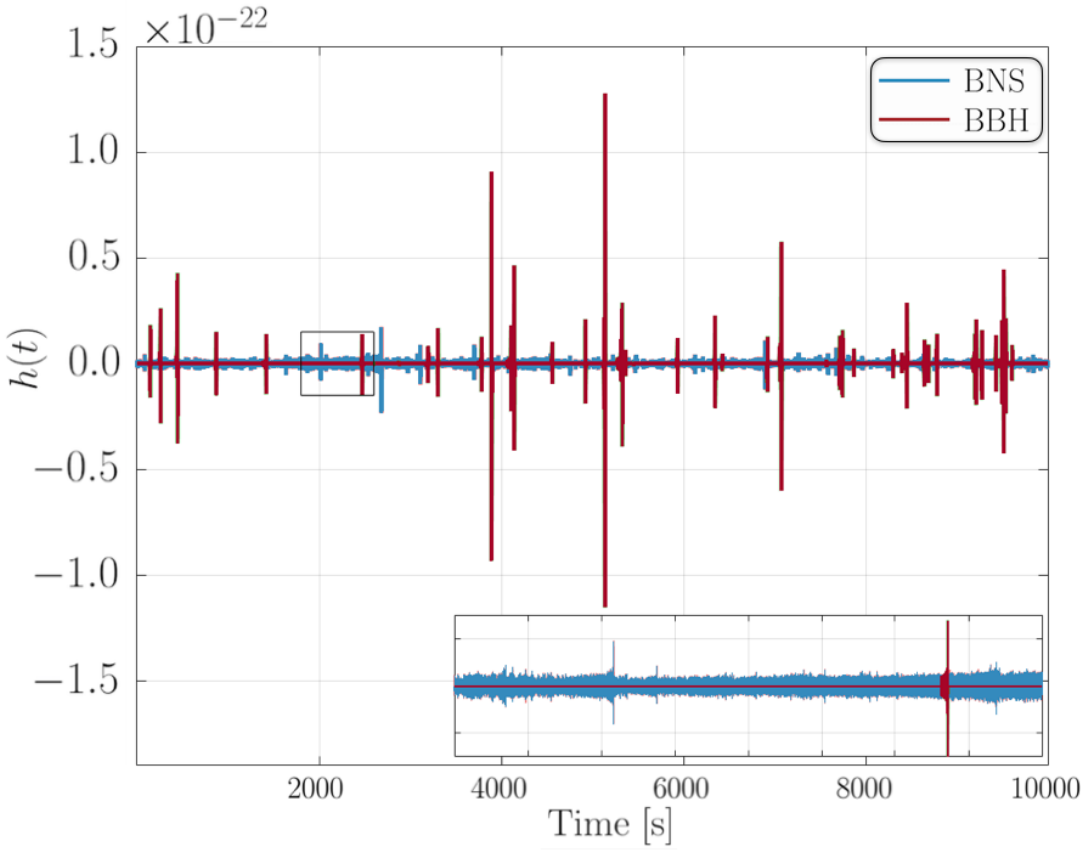

Figure 4.2: The relatively high rate of BNS mergers to BBH mergers gives rise to a background of "hisses" and "pops" colloquially referred to as the "popcorn" background. While the BNS mergers are frequent, they are of much lower strain than BBH mergers. The net result is a dull, near-constant hum of BNS mergers (shown in blue) interrupted by short and loud signals from BBH mergers shown in red.

rate, $SFR(z)$ by the relation

$$R(z) = \frac{1}{1+z}\int_{t_{\min}}^{t_{\max}} \mathrm{SFR}(z_f)P(t_d)\,dt_d \tag{4.7}$$

where z_f is the redshift at the time of formation, t_d is the time delay between formation and coalescence, $P(t_d)$ is a probability distribution of the time delay which we take to be $P(t_d) \propto t_d^{-1}$, and $t_{\max} = 13.5$ Gyr. The lower limit of the integral is $t_{\min} = 20$ Myr for BNS events and $t_{\min} = 50$ Myr for BBH events. The empirically derived fit function

for the SFR as a function of redshift is adopted from [77],

$$\mathrm{SFR}(z) = 0.015 \frac{(1+z)^{2.7}}{1 + \left(\frac{1+z}{2.9}\right)^{5.6}} \tag{4.8}$$

Although the BNS and BBH coalescence rates differ significantly, the coalescence rate as a function of redshift are similarly distributed and peaked near $z \simeq 2$ as seen in figure 4.3. We use this distribution when sampling the redshift for each of the events we create for the simulation.

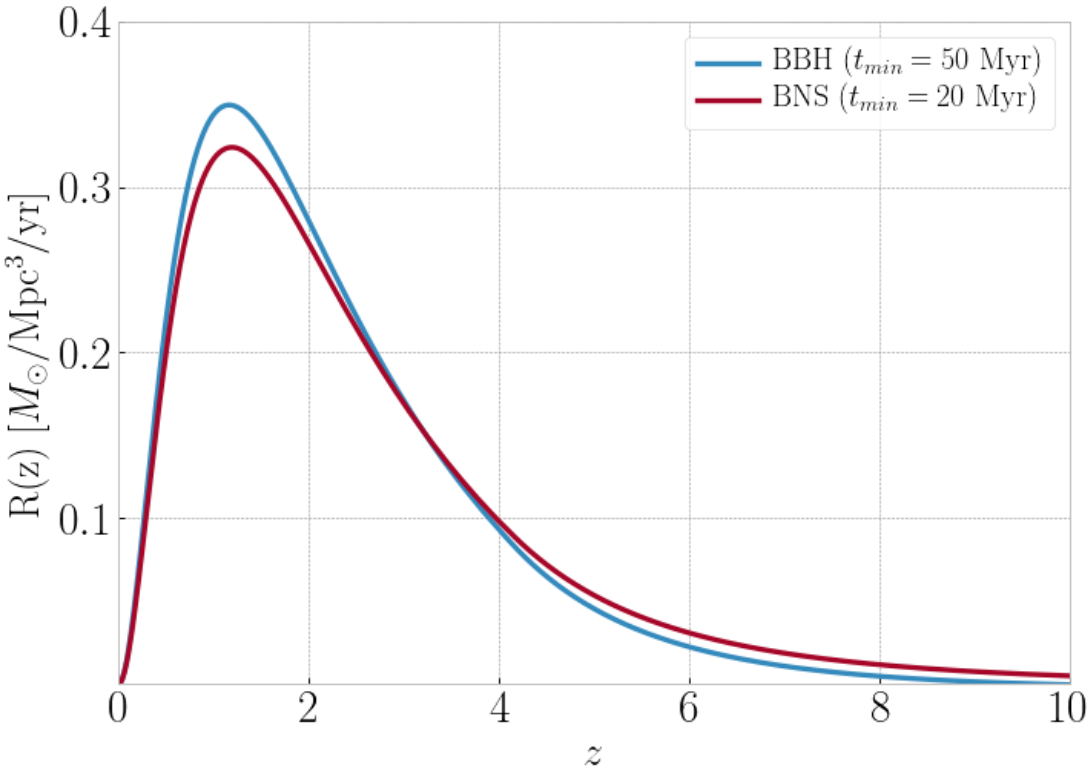

Figure 4.3: BBH and BNS coalescence rate as a function of redshift. This distribution assumes the SFR given in equation (4.8). For BBHs, we assumed a minimum delay of 50 Myr between formation and merger for all events. For BNSs, this was set to 20 Myr resulting in slightly varied coalescence distributions.

4.3.2 Mass Distributions

Next we consider the distribution of masses for BNS pairs and BBH pairs following [78, 79, 80]. For binary black holes, we draw the primary mass from a probability distribution $p(m_1) \sim m_1^{-2.35}$ where $5M_\odot \leq m_1 \leq 50M_\odot$. The secondary mass is drawn from a uniform distribution on $[5, 50]M_\odot$ subject to the condition $m_2 < m_1$. This condition is a restatement of the familiar symmetric mass distribution $q = m_2/m_1$ with $q \leq 1$. Sampling from both distributions 100,000 times, we find ∼14% of events pass the criteria and are plotted in figure 4.4.

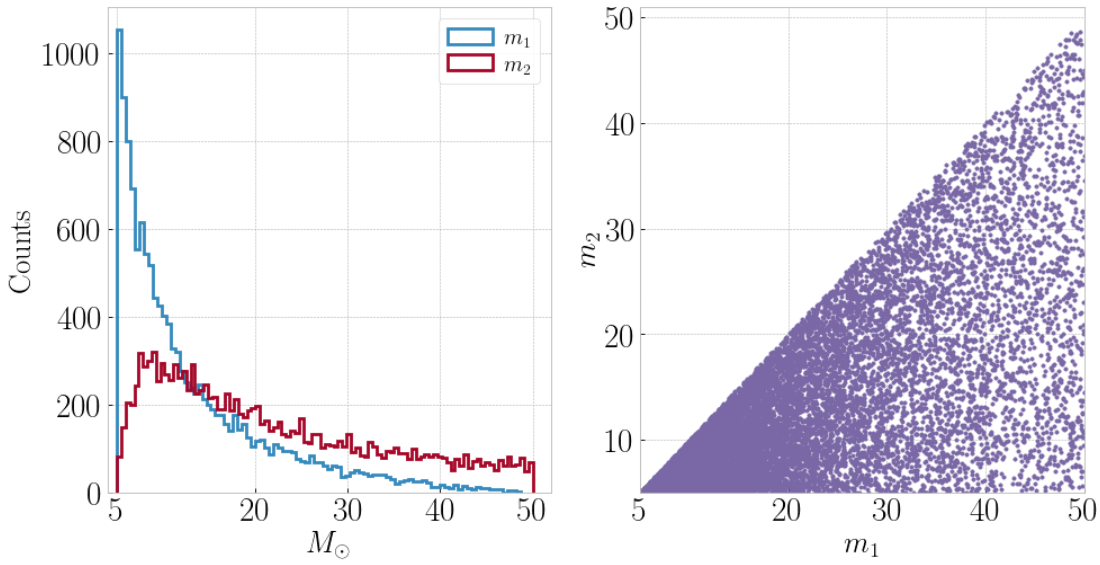

Figure 4.4: Black hole mass distributions. The lower limit on the black holes is taken to be $5M_\odot$, that is, just above the mass-gap, and the upper limit is set at $50M_\odot$. The $p(m_1) \sim m_1^{-2.35}$ behavior is evident in the histogram data collected by drawing 100,000 times and applying the appropriate selection criteria shown in the left panel. The mass-vs-mass scatter plot in the right panel shows the distribution of BBH event pairs in the simulation. The cutoff where the slope is unity is the manifestation of the condition that $m_2 < m_1$.

The neutron star mass distribution is more simplistic. Each mass is drawn from a normal distribution $\mathcal{N}(\mu, \sigma)$ with mean of $\mu = 1.4M_\odot$ and a standard deviation of $\sigma = 0.12M_\odot$ which is determined such that a mass below $1M_\odot$ or above $2M_\odot$ is a 5σ occurrence. The resulting distribution is shown in figure 4.5.

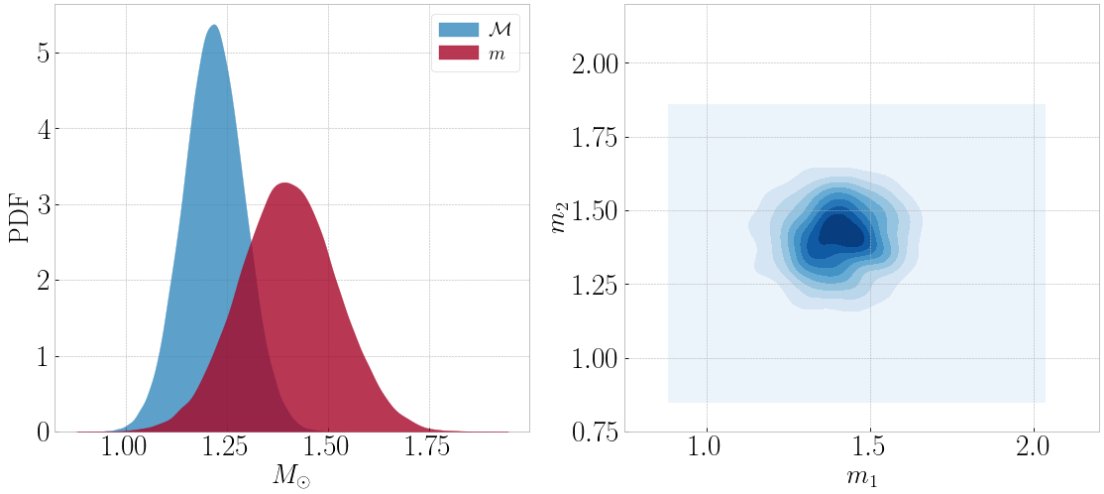

Figure 4.5: Neutron star mass distribution. Each mass is drawn from the normal distribution $\mathcal{N}(\mu = 1.4, \sigma = 0.12)$. The resulting chirp mass distribution $\mathcal{M}$, shown in blue, is therefore also a normal distribution centered at $1.21M_\odot$. The mass-vs-mass density plot is shown in the right panel.

4.3.3 Spin, Inclination Angle, Sky Location and Polarization

There are a number of parameters which are drawn from uniform distributions including the spin, inclination angle, sky localization angles and GW polarizations. The uniformity of the sky localization parameters is a restatement of the isotropy of the universe which is well supported on the scale of galaxies. We see no reason to assume that the inclination angle of the events or the polarization of events to have a preference for one value over another. There is however an apparent preference for the spin of the compact object to be close to zero (figure 4.6). This modeling is not considered in this search and we draw spins for all events from a uniform distribution on the interval $[-1, 1]$. A summary of the distributions can be found in table 4.1.

4.4 Foreground Subtraction Pipeline

Given the distributions provided above, we may create a list of CBC events with all of the parameters necessary to define the binaries. We also create two interferometer

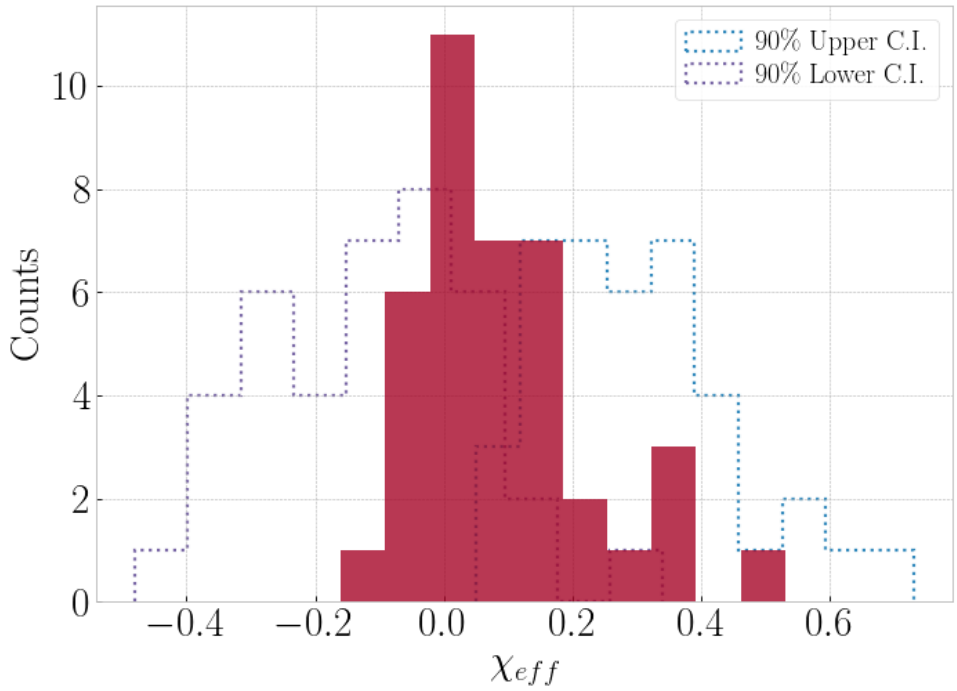

Figure 4.6: The distribution of the effective spin χ_{eff} given the 39 events in the second gravitational wave transient catalog (GWTC-2) [3]. Although this distribution does not appear to be uniform, we assume no model for the spins in this analysis and draw the effective spin parameters from a uniform distribution.

instances with the same configurations as the Hanford and Livingston detectors, but with the proposed sensitivity of Cosmic Explorer. Using the software package `Bilby` [82], we generate the signals for each of the events based on these parameters. These events are added in the time domain into each detector's strain which has been colored by white noise. A cosmological stochastic background is then added. The data is created in 4096 second segments and saved. To cut down on data generation time, a pipeline has been added to run the data generation in parallel using HTCondor. We begin the analysis with 148 frames which translates to ∼1 week of data.

Parameter	Distribution Sampling
BNS Masses	$\mathcal{N}(\mu{=}1.4, \sigma{=}0.12)$
BBH Masses	$p(m_1) \sim m_1^{-2.35}$ $p(m_2) \in [5, 50] M_\odot$
Spin (χ)	$\chi_{1,2} \in [-1, 1]$
Merger Delay	$P(\tau) = e^{-\tau/\lambda}$
Redshift	Madau-Dickinson Model (See [77, 81])
Sky Location	$\cos(\theta) \in [-1, 1],\ \phi \in [0, 2\pi]$
Polarization	$A \in [+, \times]$
BNS Fraction	94.5%
BNS Formation Rate	$p(t_d) \sim t_d^{-1}$ (20 Myr $\leq t_d \leq$ 13.5 Gyr)
BBH Fraction	5.5%
BBH Formation Rate	$p(t_d) \sim t_d^{-1}$ (50 Myr $\leq t_d \leq$ 13.5 Gyr)

Table 4.1: Distributions used to draw parameters for constructing events.

4.4.1 Parameter Estimation

For 3G detectors, parameter estimation will be incredibly challenging as the CBC signals substantially overlap. We estimate that we should observe a coalescence every ∼10 seconds, however each signal takes many *minutes* to merge after entering the range of frequencies to which the detector is sensitive (detector band). Untangling these signals successfully and uniquely will require novel, sophisticated estimation methods. A possible workaround for this signal overlap is discussed in section 4.6.1.

In this analysis, we can short-circuit this parameter estimation obstacle. Given that we know the exact parameters used to generate the mock events, we may feed these into a parameter estimation pipeline one at a time and return a companion list of the estimates of the parameters characterizing each event. This is accomplished using the `MDC_Generation` software package [83] as a wrapper for the parameter estimation code in the LSC Algorithm Library, `LAL`. The details of the parameter estimation will not be further discussed here, but the interested reader is referred to the following [84, 85, 82, 86].

4.4.2 Mask Generation

To briefly summarize, we currently have:

- Strain data realizations for two mock interferometers with the proposed Cosmic

Explorer sensitivity

- CBC injections with parameters drawn from the distributions described in table 4.1 and added into the strain data
- A GWB is injected into each mock interferometer using the Hanford-Livingston antennae patterns
- A list containing the parameter estimation for each of the events injected into the data

The process now involves generating each of the CBC events (a.k.a. "tracks") using some predetermined model and the recovered parameter estimates, and then using these tracks to mask (remove) the actual CBC tracks from the data. We choose to estimate the tracks in the frequency domain and subsequently notch them out from the cross-correlation spectrograms conveniently created as output of the isotropic SGWB cross-correlation search. We then reprocess the results of the cross correlation search having nominally removed all of the bins containing foreground CBCs.

As found in chapter 3, searches for the GWB are most sensitive at low frequencies ($\lesssim 200$ Hz). One consequence of this is that we can cut the search off above $\sim$200 Hz with a negligible impact in the search sensitivity [22]. The primary motivation for choosing to do this is because CBCs below this cutoff are largely well approximated using Newtonian dynamics [87]; the higher order relativistic terms are truncated (see figure 4.7). This tremendously simplifies the model and greatly cuts down on the computational cost of generating track estimates.

Let us now formulate the frequency evolution of a binary inspiral as a function of time in the Newtonian regime[1]. First, we will assume a circular orbit. This is appropriate since the eccentricity of a binary flattens out with $e \simeq (r/r_0)^{19/12}$ where r_0 is the radius of the binary today and r is the radius at some time in the future [19]. This flattening is quite fast and by the time a binary enters the CE detector band, the orbit will be circular to roughly one part in a million. Provided with this condition, we may invoke Kepler's law,

[1]While the relativistic waveforms are a better fit, they are computationally expensive and time prohibitive to generate with the current software package. Furthermore, at frequencies $\lesssim 100$ Hz, the relativistic dynamics are insignificant with the time and frequency resolutions being used here

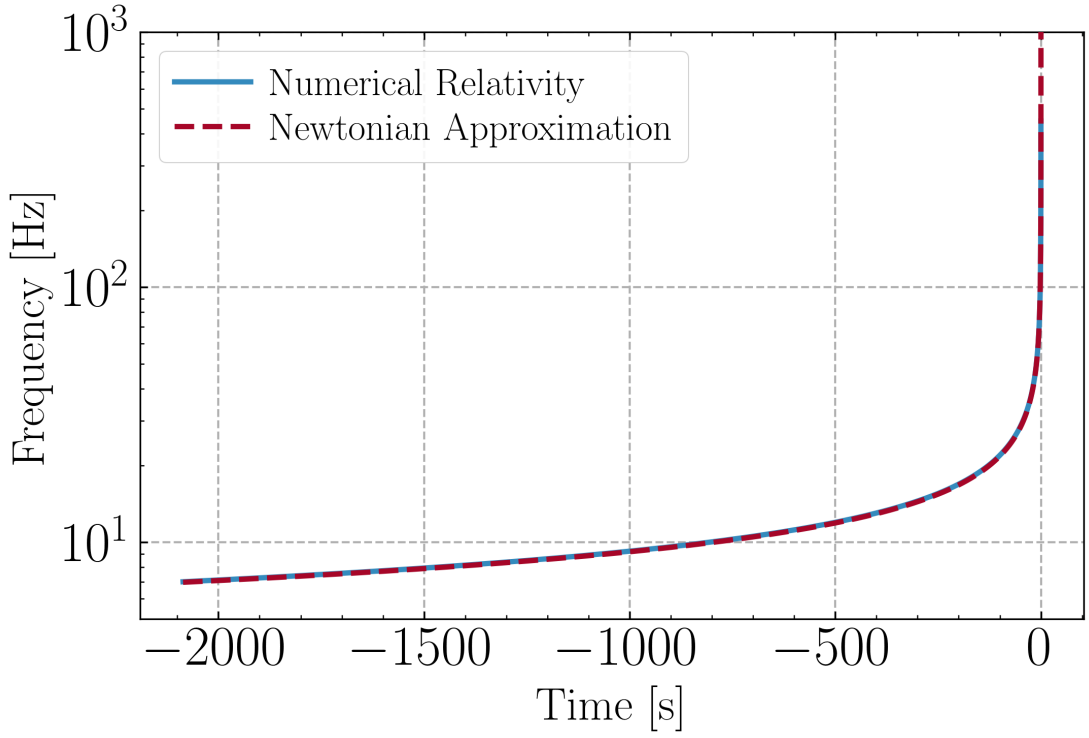

Figure 4.7: Comparison between a fully relativistic waveform and a waveform calculated with equation (4.13). Below 200 Hz, there is excellent agreement between the two methods indicating that a nonrelativistic estimate of the waveforms should be sufficient to create a reliable mask for the event. For the Newtonian estimate, the evolution of the signal is governed only by the chirp mass and the redshift.

$$\omega^2 = \frac{Gm}{r^3} \tag{4.9}$$

The energy in the orbit may then be written as

$$E_{\text{orbit}} = -\frac{Gm_1m_2}{2R} = -\left(\frac{G^2\mathcal{M}^5\omega^2}{32}\right)^{1/3} \tag{4.10}$$

where the chirp mass $\mathcal{M}$ has been inserted and the radius was substituted using equation (4.9). The total power radiated by the binary can be easily computed

$$P = \frac{32c^5}{5G}\left(\frac{G\mathcal{M}\omega}{2c^3}\right)^{10/3} \tag{4.11}$$

and we may use the fact that $P = -dE_{\rm orbit}/dt$ leading to

$$\dot{\omega} = \frac{12}{5}\left(\frac{2^5 G\mathcal{M}}{c^3}\right)^{5/3}\omega^{11/3} \tag{4.12}$$

Integrating this ODE and substituting $\omega = 2\pi f$, we find

$$f(t, t_c, \mathcal{M}) = \frac{1}{\pi}\left(\frac{5}{256}\frac{1}{(t_c - t)}\right)^{3/8}\left(\frac{G\mathcal{M}}{c^3}\right)^{-5/8} \tag{4.13}$$

where t_c is the coalescence time. We may make one final addition to this expression by accounting for the redshift of the event,

$$f(t, t_c, \mathcal{M}, z) = \frac{1}{\pi}\left(\frac{5}{256}\frac{1}{(t_c - t)}\right)^{3/8}\left(\frac{G\mathcal{M}(1+z)}{c^3}\right)^{-5/8} \tag{4.14}$$

This is the estimate for the events that we will use. Notice that the calculation of the tracks depends only upon the coalescence time, chirp mass and redshift. This model is in fact too simplistic to accurately describe all of the events generated in the mock data set. As a result, the masking technique will fail to remove the entirety of the inspiral signal from the data for each detector. This residual signal is correlated between the detectors and therefore the GWB will continue to be partly masked by this residual foreground CBC correlation. We may address this in part by increasing the width of the notching for each pixel as described below.

4.4.3 Masking Pipeline Workflow

With a model on hand, we may now begin estimating the events in the data. We will create and save a binary mask in the time-frequency domain which is zero everywhere a track is calculated to pass and unity elsewhere. Specifically, we will cycle through each frequency-domain cross-correlation file generated as output of the isotropic SGWB pipeline. For each file, we will look at each half-overlapping segment. We will then look at the CBC event list and find all of the tracks which pass through the segment. For each track, we will find the frequency at which the track enters and exits the segment.

To be conservative, we round the "entrance" frequency down to the nearest 0.25 Hz and the "exit" frequency up to the nearest 0.25 Hz. After cycling through every event within every segment of the cross-correlation file, the mask is saved. The pseudocode in algorithm 1 details the workflow.

Algorithm 1 Cosmic Explorer Event Masking Routine

```
1:  procedure GenerateMask
2:      mask ← ones
3:      for each crosscorr file
4:          for each segment in crosscorr
5:              for each event in segment
6:                  f_low ← minimum frequency of track in segment
7:                  f_low ← round_down(f_low)
8:                  f_high ← maximum frequency of track in segment
9:                  f_high ← round_up(f_high)
10:                 mask[f_low:f_high] ← 0
11:         return mask
```

4.5 Simulation and Masking Performance

We created 148 frame files at 4096 seconds each and injected each frame with CBC events according to the distributions described in this chapter. Of the 49,051 events generated, 5.5% are BBHs, 94.5% are BNSs and we exclude NSBH events as there is only one event candidate to date. A SGWB is injected into the data from each detector at the level $\log_{10}(\Omega_0) = -10.4$, that is,

$$S_{GW}(f) = \frac{3H_0^2}{10\pi^2}\frac{\Omega_{GW}(f)}{f^3} = \frac{3H_0^2}{10\pi^2}\left(\frac{10^{-10.4}}{f^3}\right)\left(\frac{f}{f_{\text{ref}}}\right)^{\alpha} \tag{4.15}$$

To help visualize the process, we look at the injections in a single 4096 second segment in figure 4.8. We find there to be 12 BBH mergers and 381 BNS mergers, a fraction of ∼3% BBH signals. The average time between coalescences is just over 10 seconds in agreement with the current rate estimates and equation (4.6). It should be emphasized that the data shown in figure 4.8 is not binned. Rather this is simply the plot of the tracks in the $f-t$ space. Since we use half-overlapping 4 second segments and a frequency resolution of $\Delta f = 1/\Delta t = 1/4$ Hz, there is considerable overlap of the

	Estimated PSD			True PSD		
	$S/10^{-11}$	$\sigma/10^{-13}$	SNR	$S/10^{-11}$	$\sigma/10^{-13}$	SNR
Noise + SGWB	4.002	6.565	60.953	1.827	6.316	28.969
Noise + SGWB + CBC	99.480	7.434	1338.254	121.425	6.316	1922.504
Noise + SGWB + CBC (Notched)	4.003	14.243	28.106	6.060	12.345	49.090
Noise + CBC	99.384	7.377	1347.195	119.357	6.316	1889.761
Noise + CBC (Notched)	3.490	14.139	24.822	5.596	12.345	45.329
Noise + SGWB ($f_{low} = 20$ Hz)	4.084	13.674	29.865			

Table 4.2: Results of the CE notching algorithm and comparisons between different data sets and using the true Cosmic Explorer PSD versus the estimated PSD using the adjacent segments. The SGWB was injected at the level $\log_{10}(\Omega_0)$=-10.4 giving $S = 10^{-10.4} \sim 4 \times 10^{-11}$. After injecting the CBCs, the foreground clearly dominated the analysis. Upon notching, the CBC foreground is reduced by a factor of $\sim$21 and the SGWB is detectable at the level of $\sim 5.6 \times 10^{-11}$. The loss of data reduces the sensitivity by a factor of $\sim$2.

event signals, particularly so at low frequencies.

The isotropic cross-correlation analysis was then performed and the $f-t$ maps of the cross correlation are calculated vis-à-vis equation (3.29). Equipped with the maps, the Newtonian model was then used to estimate tracks and create a mask of the events. When generating the bins that must be notched, we include one bin backwards in time for each calculated frequency. The Netwonian estimate is somewhat too simplistic to capture the full dynamics of the relativistic waveform, however the extra "width" to each track estimate is better able to remove the event. The masks were then applied to the $f-t$ maps and the updated results of the cross-correlation search were recorded. The results of applying the calculated mask to the raw data are shown in figure 4.9. It can be seen that the power in the bins surrounding the visible, high power tracks is underestimated. This is a result of using multiple (here we use 11) adjacent segments on either side of a central bin to estimate the noise power in the bin. When a pixel with high power in encountered, such as those where a high-mass BBH event passes, the estimate of the power across the bins is driven up and therefore the pixels adjacent to high power bins are underestimated.

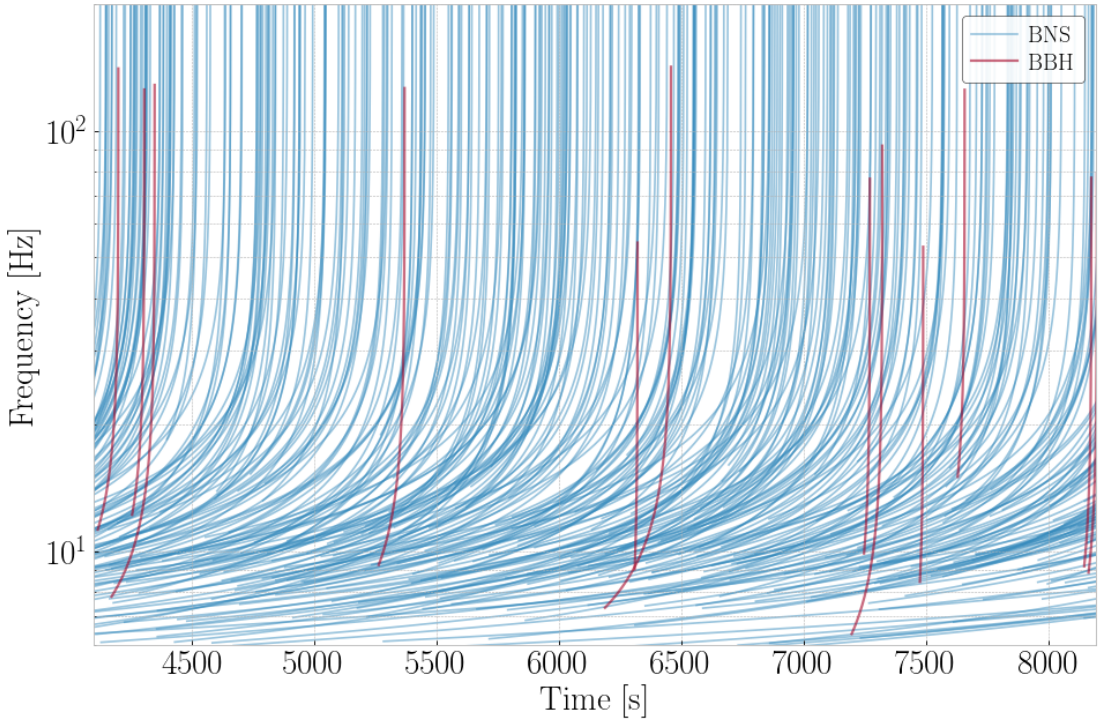

Figure 4.8: Mock CBC signals in a 4096 s frame viewed in the $f-t$ space. In this map, approximately 97% of all signals come from BNS events. On average, there is a coalescence every ∼10 seconds in agreement with the O2 rate estimates.

4.5.1 Efficiency

Figure 4.10 shows the percentage of pixels removed as a function of frequency in the entire analysis. It is immediately evident that the sensitivity below ∼20 Hz is severely compromised. Referring back to table 4.2, we see that the sensitivity achievable with CE for search for the SGWB absent CBCs is $\sigma = 6.65 \times 10^{-13}$. The current foreground subtraction routine in the ideal case of known PSDs for each segment shows that the SGWB could be detectable at 5.6×10^{-11} with 3G detectors. The foreground subtraction results show that the CBC foreground has been removed by a factor of ∼21, however we have paid the price by inadvertently removing much of the frequency band that contributes the most sensitivity to the search; a trade-off that is largely inevitable. Also we note that in the weak signal limit, we would expect *(Noise + SGWB) +*

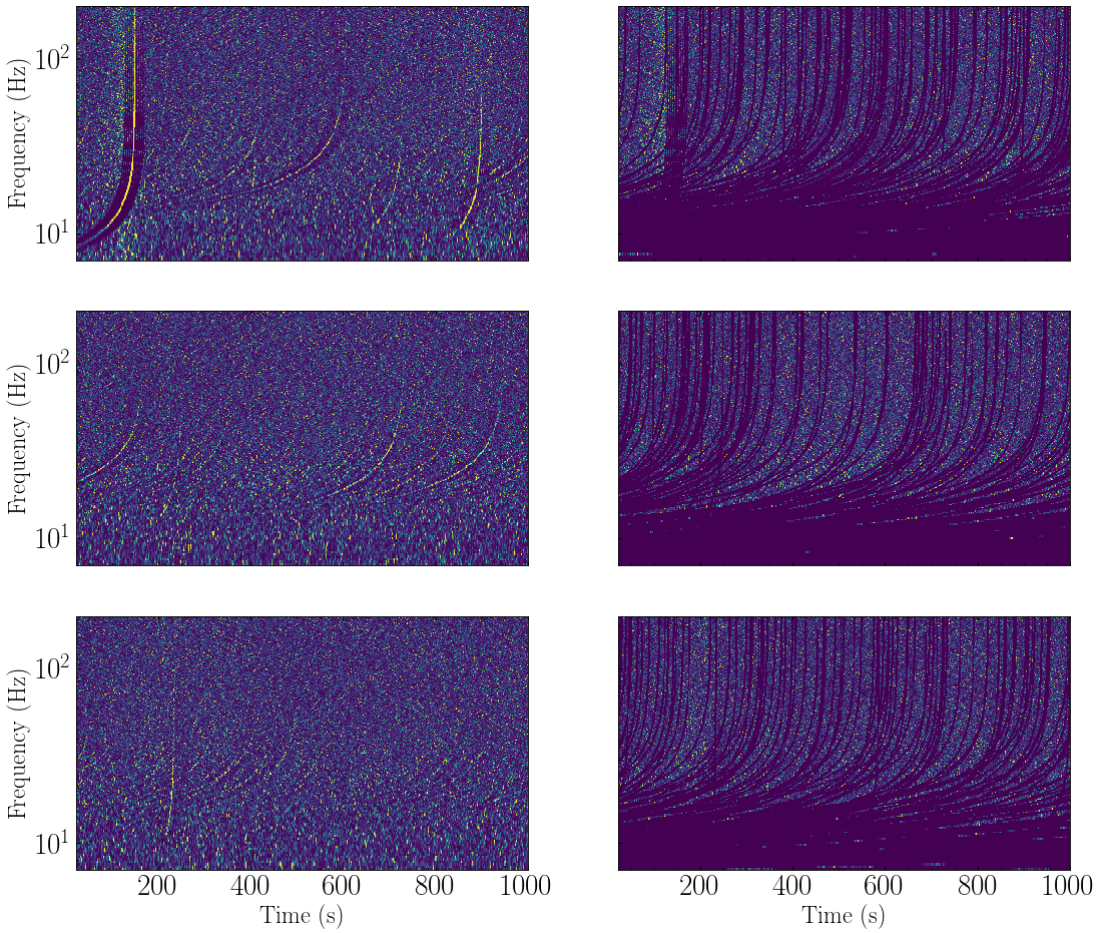

Figure 4.9: Before and after masking cross-correlation $f-t$ maps. Each row is a separate 1024 s sample cross correlation spectrogram. The left column is the raw data and the right column is the same data after generating and applying the mask. In this instance we can see that all visible tracks are successfully removed by the masking procedure. Although only a few events, mostly binary black hole mergers, are visible, there are roughly 100 events in each 1000s spectrogram. The cross-correlation data below ∼15 Hz is largely removed. Lastly, we can note the dark "fringes" surrounding the visible events in the raw spectrograms of the left column. This occurs since we use many adjacent segments to estimate the power in each bin. Therefore the higher the power in a bin, the more the power in the adjacent segments is underestimated.

(Noise + CBC, Notched) = (Noise + CBC + SGWB, Notched). However, this is not observed, possibly because the signals are strong and the weak-signal limit is not a valid

approximation.

We have also shown the results of the masking routine when no SGWB is present. From this we can conclude that there are still pixels containing tracks that were not removed in the analysis and the sensitivity is still fundamentally limited by the foreground, albeit to a much lesser extent. While the background is not fully recovered in this simple simulation, there are many performance upgrades that can be made to the model and simulation to aid in a more effective recovery process in future iterations.

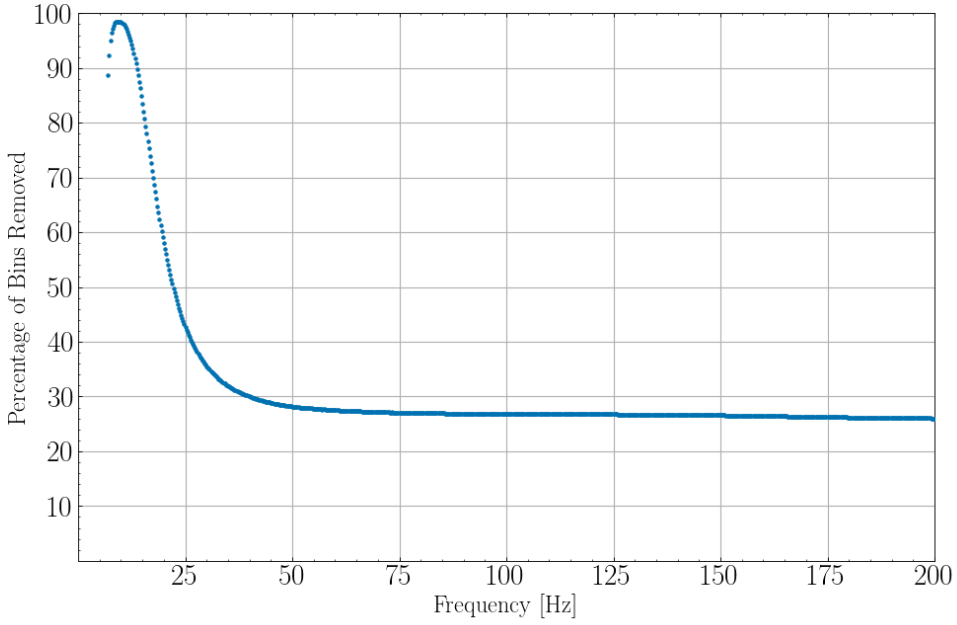

Figure 4.10: Percentage of frequency bins removed after applying the CBC foreground masks. As can be seen, the low frequency content is strongly affected by the masking. This occurs because the BNS and BBH inspirals are very nearly flat in ft-space at low frequencies. The overlapping nature of the signals in the detector means that gradually sweeping, long notches are present at low frequencies in the spectrum resulting in a steep loss in sensitivity below $\sim$25 Hz. The peak in the distribution is as consequence of the track generation technique. The events are limited to 2048 seconds. Any event which surpasses this duration is trimmed resulting in too few notches at low frequencies.

4.6 Model and Process Improvements

The pipeline presented here provides a benchmark for future work that aims to subtract the CBC foreground of 3G detectors. We implemented a first order frequency-domain, phase-agnostic approach in order to maintain simplicity and computational efficiency. While the results are encouraging, they also demonstrate a need for more subtle calculations in order to preserve the sensitivity in the low frequency regime and fully capture the dynamics of the BBH/BNS mergers. Without leaving the time-frequency space, there are a number of improvements which can be made to enhance the performance of this algorithm.

First, the search can be extended to higher frequencies to account for models with $\alpha \sim 3$ and the waveforms can be promoted to post-Newtonian models that make use of all of the recovered parameters thereby better fitting the waveforms.

Second, a wavelet decomposition could be used. It has been shown that much of the data below 20 Hz has been removed and because the GWB analyses are most sensitive at low frequencies, this is a considerable loss and limitation. If the data were decomposed using wavelets, it should be possible to have a higher frequency resolution at low frequencies, and a low frequency resolution at high frequencies. In other words, the bins at low frequencies would be wide (in time) and short (in frequency) and vise versa at high frequencies. This decomposition more closely resembles the chirps themselves and may result in much less data being removed by the analysis.

Third, look-up tables could be created for the waveforms. Currently, as seen in algorithm 1, each track is recalculated in its entirety for each half overlapping segment through which it passes. Since the events are $\mathcal{O}(10^3)$ seconds long, this means calculating the same event 500 times or more. For the first order Newtonian model this is not terribly expensive, however for post-Newtonian models this is a clear place for process improvements.

As a final note, there are a number of different approaches being explored for detecting the SGWB in the presence of an astrophysical foreground for 3G detectors such as subtraction-noise projection methods [88] and Bayesian analyses [89, 90]. These model are more complex and computationally expensive, however aspects of them could be ported over to this simulation to improve the estimation of the foreground signals and

therefore the recovery of the background.

4.6.1 Toy Model: Neural Network Image Segmentation

In addition to the methods above, there exists a separate approach that has the benefit of being fast[2] and perhaps most significantly, avoiding the highly involved and computationally expensive work of estimating the parameters of an unknown number of overlapping waveforms. To wit, convolutional neural network image segmentation architectures, such as the U-Net used in chapter 5, may prove useful. It has been shown that neural networks are able to capture the dynamics of CBC signals [91, 92, 93] and therefore once trained, a segmentation network may be an efficient method of masking the foreground.

Here we show the proof-of-concept for network segmentation architectures in masking CBC foregrounds of 3G detectors. We create spectrograms colored by Gaussian noise and then inject CBCs from the same distributions as above but at an SNR$\sim$10. Injection and detector parameters are identical to those used above. We will however split the data into 128 s chunks instead of 4096 s to make the training faster.

The network is a U-Net architecture. The downsampling block consists of four convolutions and a max pooling (*conv, conv, conv, max_pool, conv*), each using 3×3 kernels and Gaussian error linear unit (gelu) activation functions [94]. We use the ADAM optimizer [95] and a binary crossentropy loss function since the network output should be either 1 or 0. The full code, including the data presented here and trained network weights, is available in the `CosmicExplorer_MDC` repository [96].

To train and test the network, we create 100 input/target pairs where the training data consists of the standardized spectrogram and the target is the true binary mask. We will use 80% of the data for training. The results after 1000 epochs of training can be seen in figure 4.11 below. The performance of the network is somewhat conservative. That is, $\lesssim 1\%$ of the pixels containing events remain after applying the mask and $\lesssim 4.5\%$ of pixels are removed which contain no background.

[2] Assuming the model is trained and robust to diverse waveform dynamics

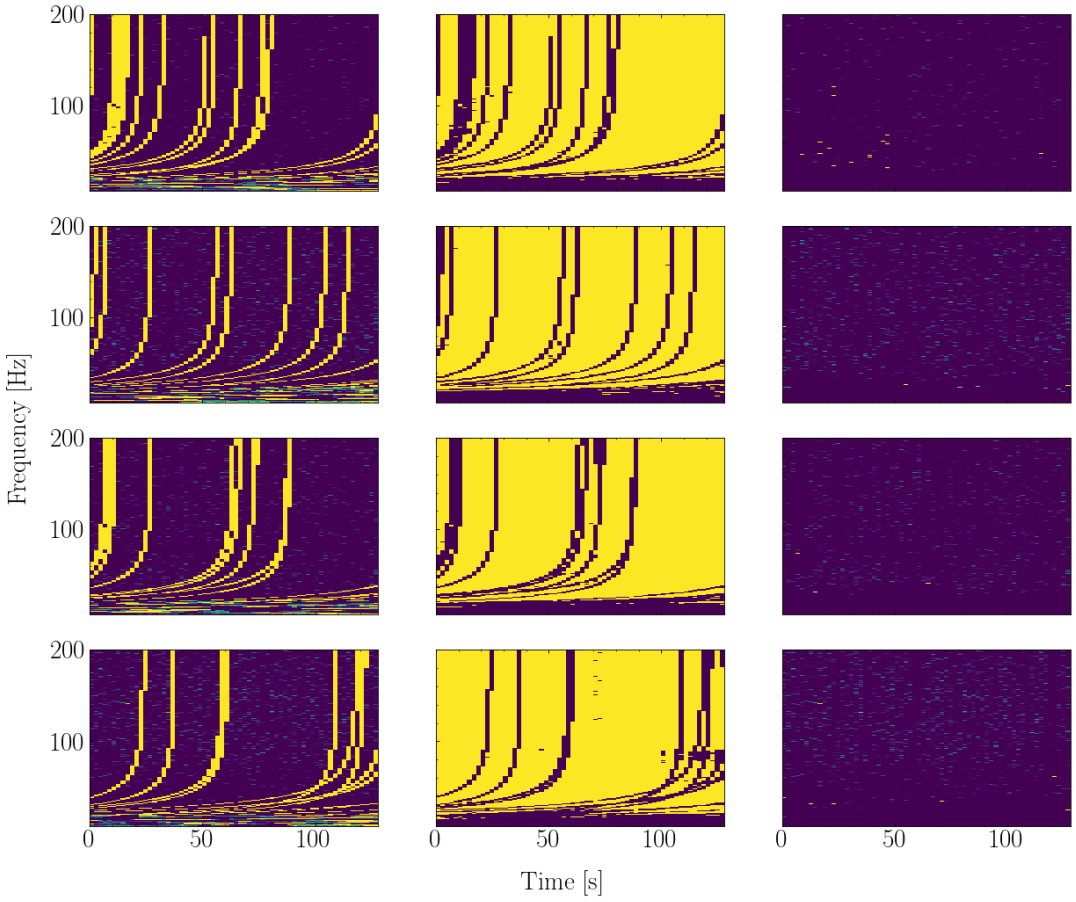

Figure 4.11: Sample results from the segmentation network. The left column is the original input training data. The middle column is the mask predicted by the network and the rightmost column is the resulting notched spectrogram. The network leaves $\lesssim 1\%$ of the pixels containing events and removes $\lesssim 4.5\%$ of the pixels which contain no events.

Chapter 5

Deep Learning Regression Methods

With the recent advent of gravitational wave astronomy, techniques to extend the reach of gravitational wave detectors are desired. In addition to the stellar-mass black hole and neutron star mergers already detected, many more are likely to be below the surface of the noise, available for detection if the noise is reduced enough [2, 3, 14, 97, 98]. In this chapter we present a machine learning framework and software pipeline (`DeepClean`) which uses gravitational wave detector data and data from on-site sensors monitoring the instrument to reduce the noise in the time-series due to instrumental artifacts and environmental contamination. While analytic filters can and are used [38, 99], deep learning algorithms provide the flexibility to capture unidentified and unknown couplings of arbitrary order whether stationary or non-stationary. Therefore, neural networks provide the possibility of constructing a "one-shot" all-inclusive filtering approach. This strategy retains the benefit of instructing hardware developments and layouts based on observed noise couplings provided with analytic methods. The GW signal fidelity, preservation and the algorithm's ability to efficiently remove noise with no unintended effects on gravitational-wave signals is also addressed through software signal injection and parameter estimation of the recovered signal.

5.1 Foundations of Machine Learning

While a full treatment of machine learning algorithms (MLAs) is beyond the scope of this chapter, some ideas will be developed as they will be used in what follows (The interested reader is encouraged to look here [100]). The fundamental idea behind supervised machine learning is to minimize the error signal that describes the difference between the true, known answer and the current estimate. This requires a "cost" or "loss" function to quantify the error signal as well as a method of updating the network weights that determine the current estimate. To understand the cost function, consider the actual state of a system to be represented by an arbitrary $2-$dimensional vector $\mathbf{S}$ and the current estimate of the state as determined by the MLA to be the $2-$dimensional vector $\mathbf{C}$. We can then construct an "error surface" $\mathbf{J}$ using $\mathbf{J} = (\mathbf{S} - \mathbf{C})^2$ which is represented in figure 5.1. Given a random initialization of the MLA estimate $\mathbf{C}_0$, we begin on a random part of the error surface, however the objective remains the same - update the network estimate to reduce the error of the true state with the network state estimate. Once the error is minimized, corresponding to the lowest point vertically on the error surface, we then say that we have reached "convergence."

To formalize this concept, assume that the state estimate of the network is determined by a set of weights $\mathbf{w}$. Generically, the error can be stated as

$$\mathbf{J}(w) = \sum_i f(\mathbf{C}(w_i) - \mathbf{S}_{(i)}) \tag{5.1}$$

where $f()$ is a smooth, differentiable function, $\mathbf{J}$ is known as the cost function eluded to above and $\mathbf{S}_{(i)}$ in the i^{th} sample of the system state. Since $\mathbf{J}$ describes the current error, then $\nabla\mathbf{J}$ describes the direction in which the error is increasing the fastest. Since we would like to decrease the error, we can simply subtract this term from our current weights to provide a new estimate,

$$\mathbf{w} := \mathbf{w} - \eta\nabla_w\mathbf{J}(w) \tag{5.2}$$

Here η, known as the "learning rate," is a hyperparameter that tells the weights how much they should update. This method of updating the weights based on the gradient of the error is known as "stochastic gradient descent" and forms the basis of a very large

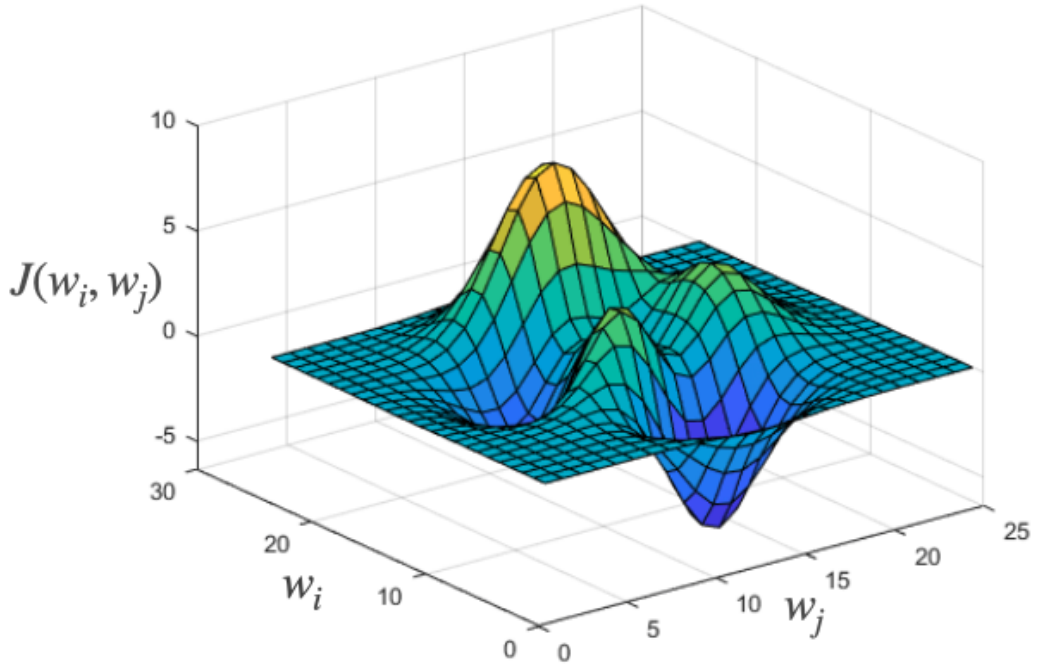

Figure 5.1: Given a current estimate of the state by the neural network and the true state, we can construct an "error surface" describing the distance from the true state to the estimate.

category of weight optimization procedures. Notice that if the learning rate is zero, the weights never update and therefore the network state never changes. Alternatively, if the gradient of the cost function vanishes, the weights again stop updating. In this second case however, we note that zero gradient means that the error has stopped decreasing. In other words, the network has found a minimum of the error surface and has converged there. There is no requirement that the converged upon minimum is the global minimum.

5.1.1 Feedforward Networks

The most basic type of neural network is called a "feed-forward" network. In this setup, the input is passed forward through the network weights to produce an output estimate.

This differs sharply from recurrent neural networks (RNN) for example which feed the output estimates back into the network as future input. Suppose we have a system state $\mathbf{x}$, a linear single layer feed-forward network $\hat{F}(w)$ and an output estimate $\mathbf{y}$. A linear network would simply be written as

$$\mathbf{y} = \hat{F}(w)\mathbf{x} = \begin{pmatrix} w_{00} & w_{01} & \dots w_{0n} \\ \vdots & \ddots & \vdots \\ w_{m0} & \dots & w_{mn} \end{pmatrix} \begin{pmatrix} x_0 \\ \vdots \\ x_n \end{pmatrix} \tag{5.3}$$

Here the output $\mathbf{y}$ would be $m-$dimensional. As we encounter many nonlinear systems, we would like the MLAs to be able to correctly model these processes as well. To make the network nonlinear, we can apply a nonlinear "activation function" $g(\cdot)$ to $\mathbf{y}$ elementwise,

$$\mathbf{y}' = g(\mathbf{y}) = g(\hat{F}(w)\mathbf{x}) \tag{5.4}$$

If we again consider the true state of the system as given by $\mathbf{S}$, we can calculate $\mathbf{J}(\mathbf{y}', \mathbf{S})$ and update the network weights accordingly. We then get a new sample of the true state, $\mathbf{S}'$ and repeat the process until convergence is reached or the number of desired iterations is achieved.

The extension to "deep" neural networks (DNNs) is straightforward. We may construct a series of matrices of weights to transform the network state into an estimate of the true state

$$\mathbf{y} = g_{(n-1)}[\hat{F}_{n-1}(w^{(n-1)})] \cdots g_0[\hat{F}_0(w^{(0)})]\mathbf{x} \tag{5.5}$$

where the $g_{(n-1)}(\cdot)$ is the n^{th} elementwise nonlinear function, $\hat{F}_{(n-1)}()$ is the n^{th} layer of the network, and $w^{(n-1)}$ is the n^{th} set of weights of the system. Once the output state is calculated, it is compared to the true state of the system and a metric of the error is again calculated. This time however, the chain rule (as opposed to a single gradient) must be invoked in order to pass the error signal "backwards" through the entire network and update each of the weights in all n layers. This process is known as back-propagation and will not be discussed further.

Equipped with a network architecture, true state samples, activation functions, and

a loss function, we may proceed to iterate input through the network until the cost function vanishes. While this description is quite simple and there are many other types of network, they all have the same general ingredients: an architecture which contains one or more (nonlinear) layers each with one or more weights, a metric to determine the error of the estimate and a method by which the weights are updated.

In this chapter, we will consider convolutional neural networks (CNNs) which use many kernels with dimensions less than that of the input. These kernels are convolved with the input and each kernel produces an output known as a feature map which is reduced in shape relative to the original input. In subsequent layers, each feature map is convolved with more kernels to produce another set of feature maps which are again smaller that in the previous layer (see figure 5.2). Since here we want the output to be of the same dimension as the input, we will apply de-convolution operations subsequent to the convolution steps until the feature maps are the same size as the original input as shown in the architecture of figure 5.6. These maps are combined and produce the final estimate of the true state.

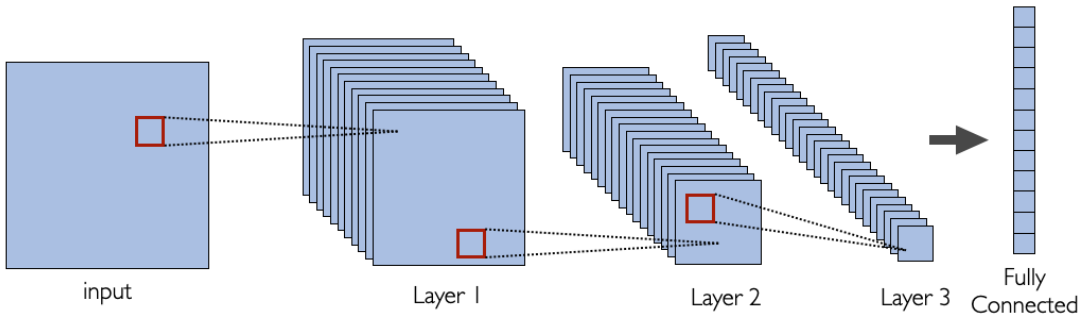

Figure 5.2: Schematic of a convolutional neural network. The red squares correspond to a filter kernel. These kernels are convolved with the input and with the feature maps of each layer. After the final convolution, the feature maps are 'unraveled' to create a feature array known as the fully-connected output. The network then uses the values in the fully connected layer to score the input.

5.2 Formalism and Loss Function for LIGO Data

The gravitational wave strain signal is collected as a time series where the amplitude is determined by the intensity of the laser light at the output photo diodes (PD) of the detector due to a phase mismatch of the recombined beams. The misalignment of the phase moves the photo diodes away from the dark interference fringe. The amount of physical displacement is related to the amount of light falling on the PDs and is determined through manual calibration methods. The strain is the first order expansion of the metric as measured by the detector

$$h(t) = h_{\mu\nu}(t)D^{\mu\nu} \tag{5.6}$$

where $D^{\mu\nu}$ is known as the detector tensor and it encodes the geometry of the detector's sensitivity. Unfortunately, the strain $h(t)$ that is actually measured is not simply the influence of gravitational waves $s(t)$. It also contains the noise sources $n(t)$ which couple into the detector. These noise sources can further be subdivided into to categories

$$n(t) = n_w(t) + n_{nw}(t) \tag{5.7}$$

where $n_w(t)$ represents the "witnessed" noise sources which have a direct auxiliary monitor and we expect to be able to remove from the output strain data, and $n_{nw}(t)$ represents fundamental "not witnessed" noise sources we do not expect to be able to remove such as quantum noise and thermal noise.

The objective is to design a network (a CNN) which takes in multiple auxiliary channel sensor time series, finds couplings between the target output and the sensor data, and subtracts the sensors' contribution to the noise from the target output. Therefore, the network we seek to create aims to provide an estimate of the witnessed noise via minimization of the cost function, and not to estimate GW signals themselves.

The auxiliary sensor data can contribute to the total noise floor as a weighted and phase shifted sum of linear contributions, or higher order couplings and non-stationary contributions. The network estimate $\mathcal{F}(\omega_i(t); \vec{\theta})$ described by the witness channels $\omega_i(t)$ and network parameters $\vec{\theta}$ provides a map of the witnesses to the noise and may be used to construct a filter for the data. Here the network parameters include the weights and

hyperparameters such as the activation function and optimizer. The particular network parameters $\vec{\theta}$ can be found through the minimization of a cost function J. This process is described mathematically as

$$\vec{\theta} = \text{argmin}_{\vec{\theta'}} J\left[h(t), \mathcal{F}(\omega_i(t); \vec{\theta'})\right] \tag{5.8}$$

For DeepClean, the cost function is chosen to be the weighted average amplitude spectral density (ASD) of the residual strain $r(t)$. Due to persistent high power spectral features such as calibration lines and power lines, and the steeply rising noise floor from seismic noise below ∼20 Hz (figure 5.3), the cost function has to address each frequency bin independently. If this were not the case, then the bins with the most error would always dominate the cost function and the network would not "see" the noise contributions in the more sensitive regions. This can be expressed as

$$J_{\text{ASD}} = \frac{1}{f_2 - f_1} \int_{f_1}^{f_2} W(f) \sqrt{S[r,r](f)} df \tag{5.9}$$

$$r(t) = h(t) - \mathcal{F}(\omega_i(t); \vec{\theta}) \tag{5.10}$$

where $S[r,r]$ is the power spectrum of the residual strain and the range $[f_1, f_2]$ is the frequency band of interest. The weights $W(f)$ are chosen to be the inverse of the target strain power $S[h,h]$. This provides a whitening effect demonstrated in figure 5.4 and ensures that narrow-band spectral features do not influence the estimates of the surrounding noise floor and helps the network to converge more quickly. For N frequency bins, we may write the cost function as

$$J_{\text{ASD}} = \frac{1}{N} \sum_{i=0}^{N-1} \sqrt{\frac{S[r,r][i]}{S[h,h][i]}} \tag{5.11}$$

The network performance can be improved by including the mean square error signal in the time domain summed over the number of time series samples T,

$$J_{\text{MSE}} = \frac{1}{T} \sum_{t=0}^{T-1} \left(h(t) - \mathcal{F}(\omega_i(t); \vec{\theta})\right)^2 = \frac{1}{T} \sum_{t=0}^{T-1} r[t]^2 \tag{5.12}$$

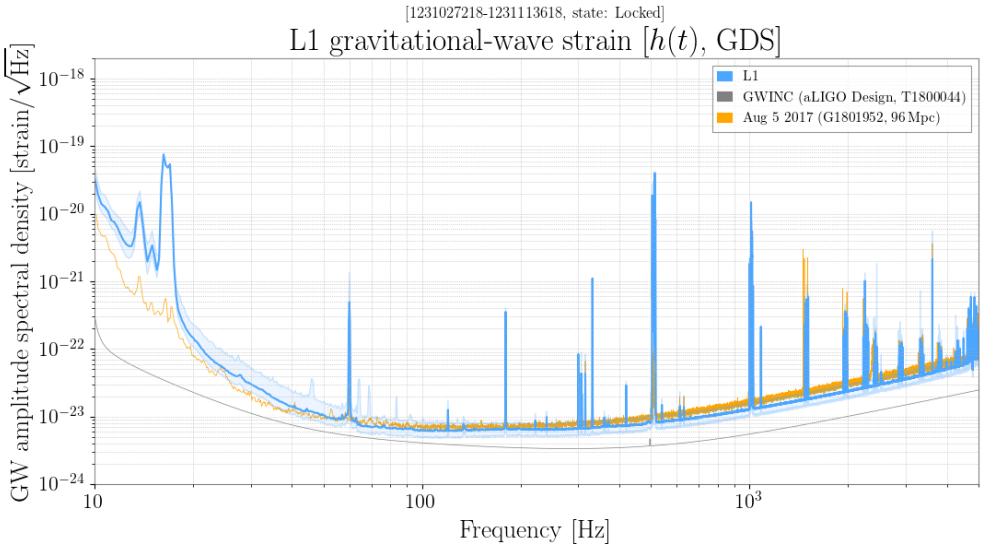

Figure 5.3: Typical strain of the LIGO Livingston detector during O3. The electrical 60Hz line and its harmonics are clearly seen as well as the calibration lines near 15Hz. One can also observe power from the oscillations of the mirrors (violin modes) slightly above 500Hz and a sharp rise at low frequencies due to seismic noise. It can be seen that there is a net difference in power of 4 orders of magnitude across all frequencies. As a result, raw strain data cannot be fed into a neural network since the spectrum is dominated by just a few frequencies and therefore the cost function and weights will only observe and update these frequencies until they are of the same order of magnitude as the rest of the spectrum. Consequently, most of the spectrum will be poorly estimated and add noise into the system. Therefore careful pre-processing steps must be taken with both the sensor and target data.

The cost function in its final form can then be given as

$$J = \beta J_{\text{ASD}} + (1 - \beta) J_{\text{MSE}} \tag{5.13}$$

where β is a hyperparameter in the range $[0, 1]$ which must be chosen by hand. We have found that training near or on narrowband features is best suited by J_{ASD} and therefore values of β near unity. For broadband noise removal however, J_{MSE} is preferable.

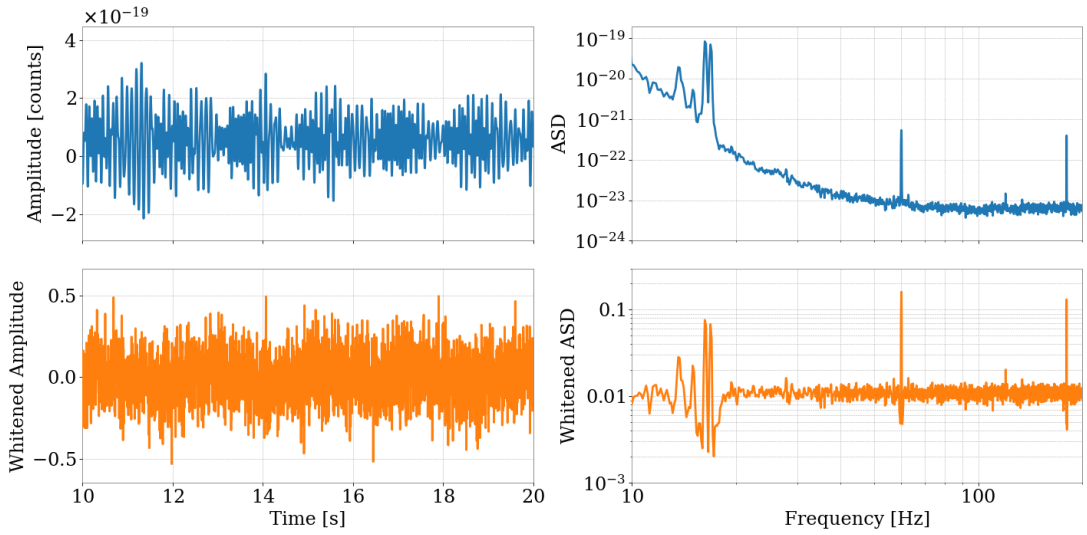

Figure 5.4: The top row shows the time series and amplitude spectral density respectively for LIGO Livingston data during O3. The bottom row is the same time series after weighting the data by the PSD of the stain data at a different observation time. While some sharp spectral features remain, the amplitudes across the broadband region in the whitened ASD vary only by approximately one order of magnitude which is sufficiently close for the CNNs to train on the entire dataset without spectral bias.

5.3 Data Selection and Preprocessing

The auxiliary channels used to regress the noise from $h(t)$ are sampled at different rates depending upon the monitor sub-system type. Seismometers measure low frequency noise and are sampled ≤ 256 Hz whereas accelerometers on the beam splitter table, and indeed DARM itself, are sampled at 16kHz. In order to maintain the proper temporal separation between channels, we resample all channels to the minimum frequency determined by the channel list and frequency band we are attempting to regress. This minimal choice reduces complexity of the data and increases convergence while decreasing training time.

After resampling, DARM is bandpassed in the time domain around the frequency band of interest using an 8^{th} order Butterworth filter with zero phase. This again aids in training time and convergence. Nonlinear couplings which may be present could

be caused by beating or coupling outside of the bandpassed region, and therefore the witness channels themselves cannot be band limited. They are however standardized

$$\omega_i(t) = \frac{\omega_i(t) - \overline{\omega_i(t)}}{\sigma_i} \tag{5.14}$$

where $\overline{\omega_i(t)}$ and σ_i are the mean and standard deviation of witness channel i respectively. The target data is standardized as well and the result is rescaled at the end of the run.

In addition to the scaling, the data set is augmented by splitting the training samples into partially overlapping 8 second segments. The network then has more opportunities to fit the various noise couplings with the trade-off being the increase in computational time. During training, the segment overlap is 7.75 s (96.875%) and during inference the overlap is 4 s (50%).

After preprocessing the data, the data is split into training and testing samples. The output of the network is the estimate of the noise in the test samples. By subtracting this estimate away from the target signal, we are able to remove the noise in this step. The workflow can be seen in figure 5.5.

5.4 Neural Network Architecture

MLAs are generally highly optimized for a particular problem; the general architecture, hyperparameters, input and output shape and type, activation functions, etc. For the regression of LIGO data, we choose a one dimensional convolutional autoencoder.

Convolutional layers consist of the input data, filter kernels with trainable weights, and a feature map. For each layer, we specify the size and number of filters to be applied to the input data. Each filter is then convolved with the input data (a sliding window dot product) and the resulting downsampled output is referred to as the feature map. Depending on the filter size and stride, the convolution may break and therefore the input data gets zero padded. For data with an initial dimension M, filters of size $K \times K$, a stride of S (number of positions the kernel is shifted relative to the input) and a padding of P, the output size O will be

$$O = \frac{M - K + 2P}{S} + 1 \tag{5.15}$$

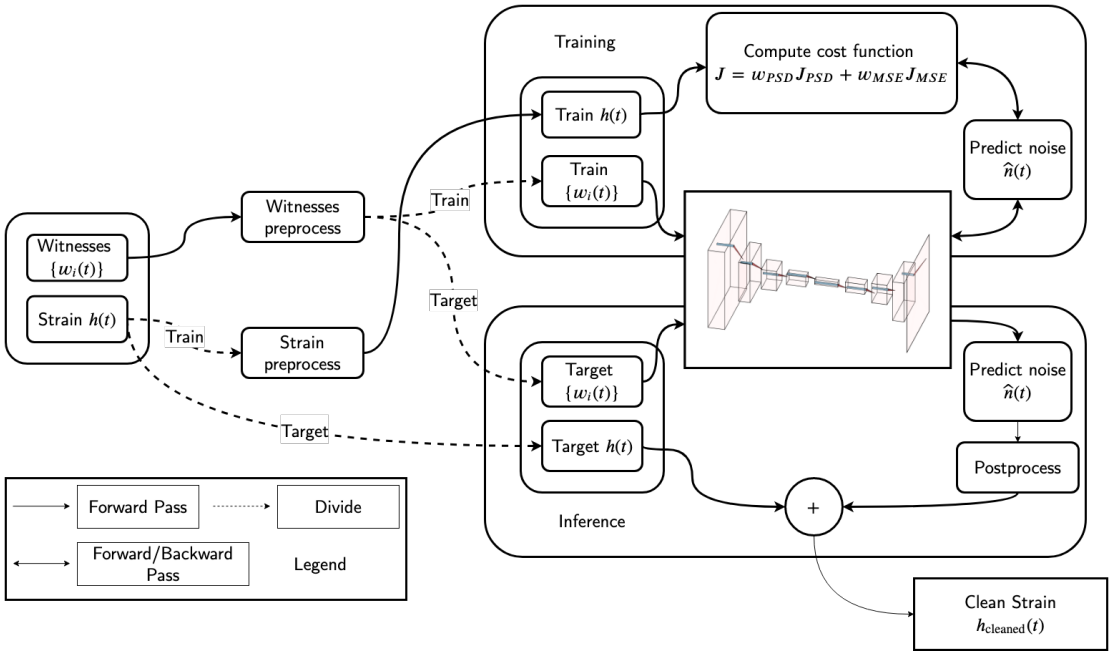

Figure 5.5: Workflow for the `DeepClean` noise subtraction pipeline. Thi image is used with permission from T. Nguyen [37].

For each convolution we will have $O \leq M$ and therefore these layers perform a high level abstraction of the data. By performing a transpose convolution, we could construct the original input data from the feature maps and the filter kernels.

It is not unusual to have hundreds of filters and feature maps for each layer, though the filters are generally 7×7 matrices or smaller. Although DeepClean is a 1-dimensional network, the filters span across the witness channels as well as across time and are therefore 2-dimensional. The filter weights are updated during the training process which should allow different filters to identify particular aspects of the data including coupling combinations which produce a good estimate of the target strain signal.

Following the convolution of the input, the current training sample (batch) is renormalized and subsequently projected onto a hyperbolic tangent curve by element-wise application of $g(x) = \tanh(x)$ as our activation function, eluded to in equation (5.5). The renormalization is helpful in training the network since the gradients between nodes does not vanish. Additionally, because we will need to calculate the gradients between

every weight across the network, this can be very intensive. Fortunately,

$$\frac{d\tanh(x)}{dx} = 1 - \tanh^2(x) \tag{5.16}$$

and so the gradient calculation is made very simple and fast with this choice. This will be discussed more in section 5.5.

The autoencoder structure is created by reducing the dimensionality of the input data. The idea is that the network will keep only the critical information; only the most important noise couplings will be retained. This speeds up the network training time, helps to remove bias and overtraining, and maximizes the use of the depth (number of layers) in the network. Since the noise regression problem requires a one-to-one input-output ratio, we must "de-convolve" the autoencoder through the transpose convolution operation. As in the case of the ordinary convolutions, we then normalize each layer and apply the hyperbolic tangent activation function to the data. The full architecture of the network is shown in figure 5.6.

5.5 Training and Inference

The DeepClean workflow is subdivided into a training pipeline and an inference pipeline. We train the network in mini-batches of 32 samples each. The loss function is then calculated, averaging over the mini-batches, using equation (5.13), and the gradient with respect to the network parameters $\vec{\theta}$, is calculated and passed backwards for weight updating. Typically, after $\sim$10 iterations over all mini-batches (epochs), the gradient descent has converged. The network is therefore given 50 epochs to converge and training is stopped at this point whether there is convergence or not. The training times vary greatly contingent upon the complexity of the noise sources in the data and the relevance of the witness channels used. We choose the ADAM optimizer for our descent [95] but use a learning rate which decays more quickly than the typical default parameters - about 10x after every 5 epochs.

In the inference step, the network uses the available witness channel information to create an estimate of the noise in the target. This is in contrast to using the data to estimate the target itself. This estimate is again processed to ensure that no estimates outside of the bandpassed region can contribute to the estimate as the network should

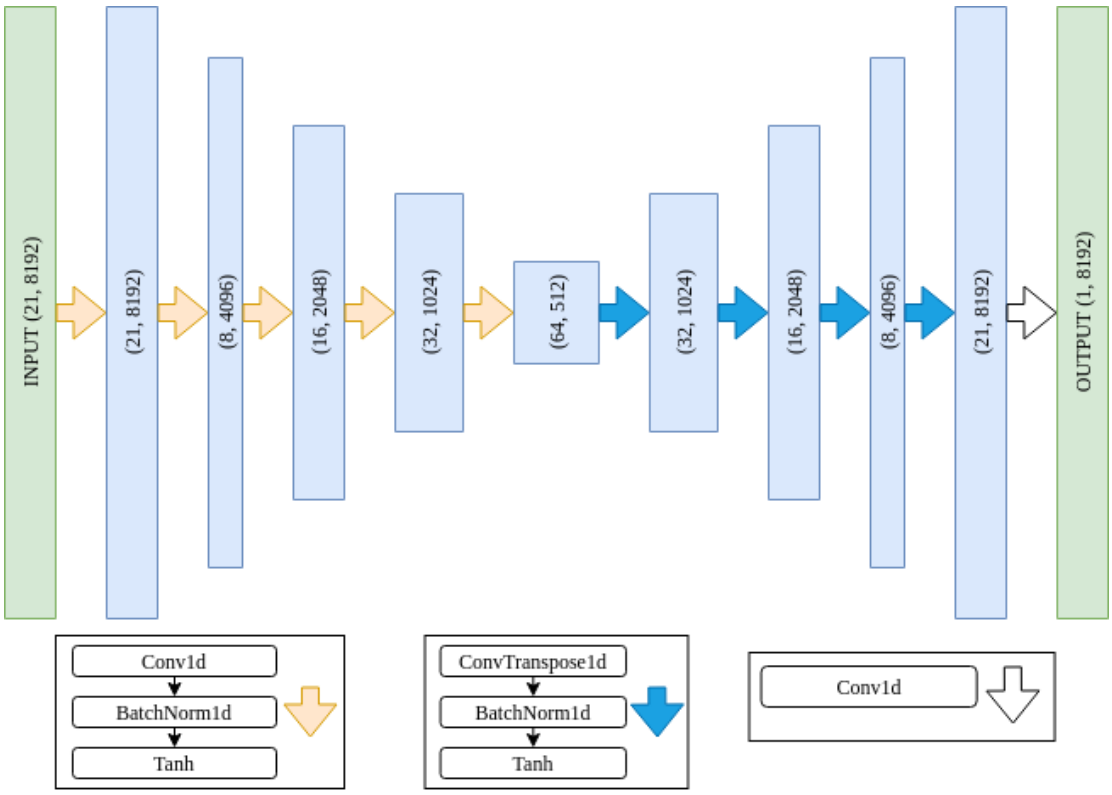

Figure 5.6: Convolutional autoencoder architecture used by `DeepClean`. The input data segments from each of the 21 witness channels are 8 s long and sampled at 1024 Hz giving an input dimension of (21, 8192). The orange arrows correspond to data dimensionality reductions through convolutions as can be seen by the shape of the data given in parenthesis on each layer. The blue arrows signify the transpose convolution operation. The final transparent arrow is the final pass from the witness channels to the target. All filter kernels are 7×7 with a stride of 2. This image is used with permission from T. Nguyen [37].

have been blind to this information during training. Finally, the noise estimate produced by the network can be subtracted from the testing gravitational-wave strain data.

H

Since not all observable noise is stationary, the network is kept as lean as possible in terms of free parameters in order to keep re-training times low. This helps to ensure that the performance of the network on the adjacent testing data is in agreement with

the results of the training. Using 8 second segments with 7.75 seconds overlapping, it takes the network roughly 6 minutes to train on 1024 second worth of data on an NVidia Titan X GPU [101, 102]. The inference of the trained network takes only a few seconds.

Because the duration of the data exceeds the training time, the network can be used off-line and on-line (real-time) subtraction. We expect that the training time will increase with $\mathcal{O}(n)$ where n is the number of auxiliary channels used during training. There is therefore a practical limit to the complexity of the noise which could be modeled or to the size of the parameter space which can be tested at one time. Determining and disregarding input data which is irrelevant to the subtraction is consequently an important part of the prepressing step.

5.6 Performance Results

The `DeepClean` noise subtraction pipeline has been applied to data from both the second and third aLIGO observing runs. In the case of the second observing run, beam jitter of the pre-stabilized laser caused by a water cooling pump on the beam splitter table dominated the contribution to the noise floor across the $\sim 50 - 1000$ Hz band. This coupling (as well as the calibration and power lines) is linear and therefore analytic Wiener filters are effective here. Neural networks should be able to capture these linear dependencies and retain the flexibility to capture nonlinear and non-stationary noises as well. This would lend the neural network filter to be a "one-shot" all-inclusive filter for LIGO data. We therefore apply our network to the O2 dataset to regress the linear noise and compare it to the optimal mean square error Wiener filter.

To perform the subtraction on the O2 dataset, we use the same start times (1164556817), sample rate (2048 Hz), duration (1024 s) and auxiliary channels (below) that were used in the Wiener filter regression [38, 103].

- H1:DCS-CALIB_STRAIN_C02 (target)
- H1:IMC-WFS_A_DC_PIT_OUT_DQ
- H1:IMC-WFS_B_DC_PIT_OUT_DQ
- H1:IMC-WFS_A_DC_YAW_OUT_DQ

- H1:IMC-WFS_B_DC_YAW_OUT_DQ
- H1:PSL-DIAG_BULLSEYE_PIT_OUT_DQ
- H1:PSL-DIAG_BULLSEYE_YAW_OUT_DQ
- H1:PSL-DIAG_BULLSEYE_WID_OUT_DQ
- H1:IMC-WFS_A_DC_PIT_OUT_DQ
- H1:IMC-WFS_B_DC_PIT_OUT_DQ
- H1:IMC-WFS_A_DC_YAW_OUT_DQ
- H1:IMC-WFS_B_DC_YAW_OUT_DQ
- H1:PEM-EY_MAINSMON_EBAY_1_DQ
- H1:PEM-EY_MAINSMON_EBAY_2_DQ
- H1:PEM-EY_MAINSMON_EBAY_3_DQ
- H1:SUS-ETMY_L3_CAL_LINE_OUT_DQ
- H1:LSC-CAL_LINE_SUM_DQ
- H1:CAL-PCALY_TX_PD_OUT_DQ
- H1:CAL-PCALX_TX_PD_OUT_DQ

The `DeepClean` network uses tanh activation functions and has no requirement to find purely linear relationships. The data was trained by first parallelizing the training/inference process to subtract the narrowband calibration lines and power lines. The outputs were combined into the preliminary data and then the broadband subtraction of the jitter noise was performed. The results of the subtraction and a comparison to the Wiener filter as seen in figure 5.7 shows the nearly identical output. The similarity of the outputs suggests not only the linearity of the channels used to regress the noise, but the robustness of the network itself to not over-train or develop spurious false relationships.

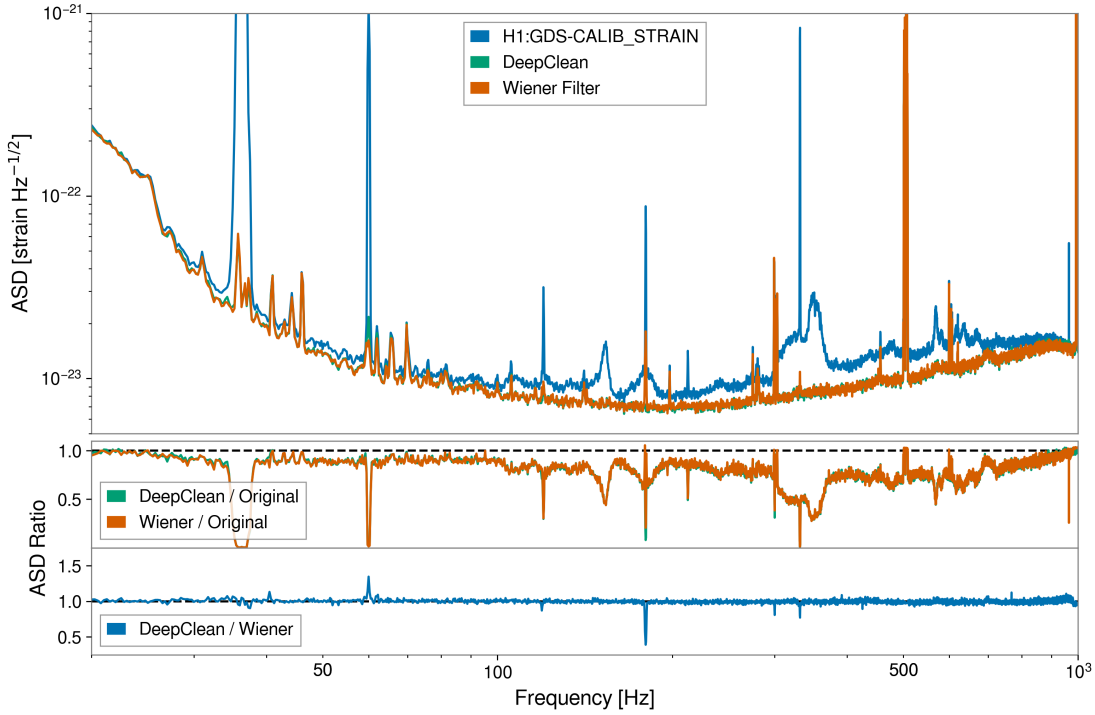

Figure 5.7: *Top:* The blue curve is the ASD of the raw strain data from LIGO Hanford during O2. The orange and green curves are the ASDs of the same stretch of data after applying the Wiener filter and DeepClean filter respectively. *Middle:* The ASD ratio of the cleaned data relative to the raw data is shown for both methods. *Bottom:* The ASD ratio of the two cleaning methods shows that the performance is nearly identical across ∼1000 Hz frequency band. This suggests that the DeepClean network has learned the optimal linear couplings of the witness channels without overfitting the data. This image is used with permission from T. Nguyen [37].

Because so much broadband noise can be removed in the O2 dataset, the effective range that the detector can measure (known as the "Binary Neutron Star Inspiral Range," equation (2.2)) increases by ∼20%. This range increase, shown in figure 5.8, translates into a detectable volume increase of ∼73%.

During O3 it was found that there was persistent, stationary power buildup around the 60 Hz peak from the power lines. These peaks were at $60 \pm 0.25n$ Hz where $n = 0, 1, 2 \ldots$ and there were no auxiliary monitors measuring these peaks. It was realized that there was slow alignment sensing and control (ASC) motion that was beating

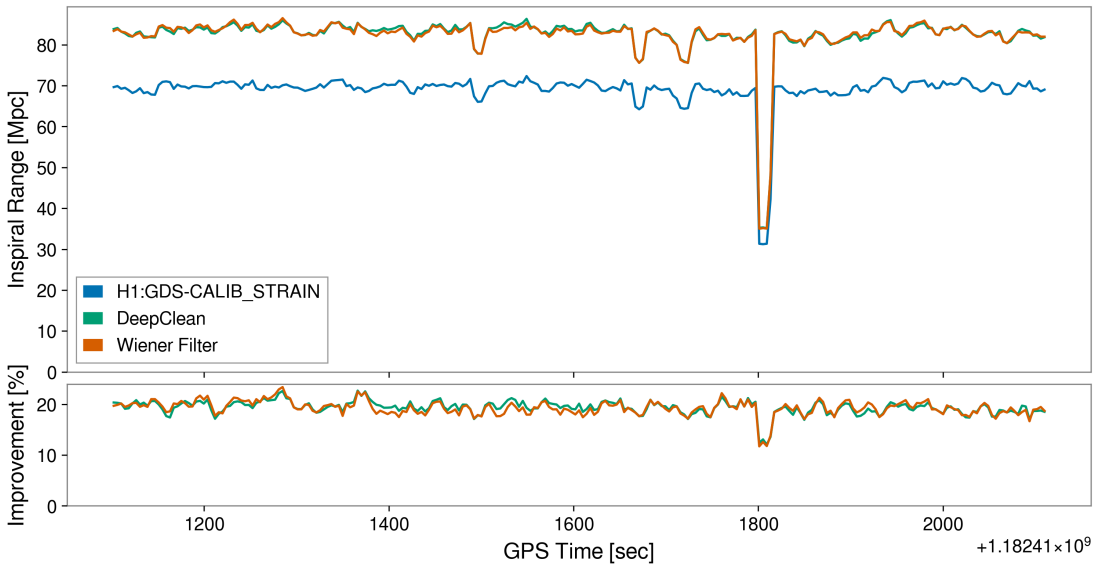

Figure 5.8: The binary neutron star inspiral range increase after cleaning the O2 dataset with a Wiener filter and with the `DeepClean` pipeline. The total range increase is ∼20% and the two filters perform within ∼1−2% of each other. This image is used with permission from T. Nguyen [37].

against the 60 Hz lines causing the sidebands. This noise was therefore a stationary, bilinear effect. Because we can model the beating between two sine waves as

$$\cos(2\pi f_1 t) + \cos(2\pi f_2 t) = 2\cos\left(2\pi\frac{f_1+f_2}{2}t\right)\cos\left(2\pi\frac{f_1-f_2}{2}t\right) \tag{5.17}$$

then the ASC modulation was at 0.5 Hz. These channels were found and could be run through the `DeepClean` network. For comparison, figure 5.9 contains the original O3 spectrum from Hanford, the data after neural network regression and the analytic second order filter results discussed in chapter 6. Based on the results, we can conclude that the network is able to remove both the linear and bilinear noise features of the data without overfitting the training data.

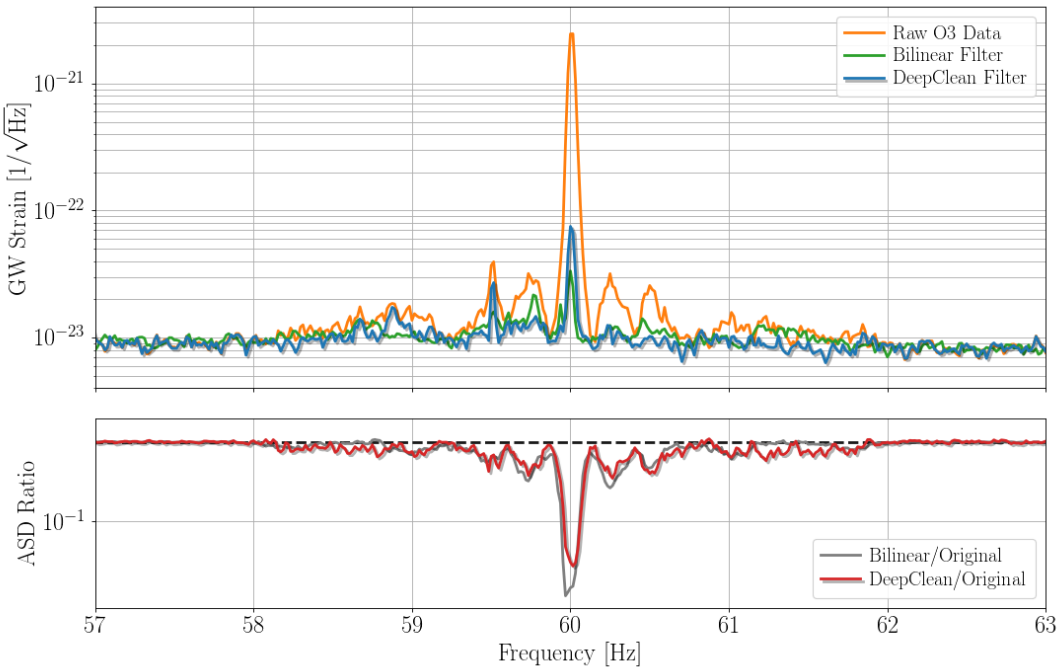

Figure 5.9: Bilinear subtraction of the 60 Hz power mains line and ASC-modulated sidebands with `DeepClean`. The top panels shows the original data (orange) as well as the results obtained after regression with the network. The second order filter results show the fidelity of the network. The bottom panel shows the ratio of the cleaned-to-original ASD. It is clear that the network does not add any power (noise) to the data and targets the broad sidebands and power line frequencies.

5.7 Safety Studies

The procedure outlined above is carried out for each gravitational wave time-series separately, i.e. the ones from the LIGO detectors in Hanford and Livingston. If the performance of the trained filters does not add noise and is either consistent with known results from analytic methods or subtracts spectral features of the target channel in a manner consistent with the features of the witness channels, then those filters are safe to apply to the strain data during the time of interest. The result of the filters is the production of a new strain time series which should have increased fidelity to the gravitational-wave strain signal incident on the instruments. One way of assessing the

ability of our method to denoise gravitational-wave time series is by invoking parameter estimation methods on a set of astrophysical signal waveforms that are injected via software and for signals in which the true astrophysical parameters are known *a priori*. The parameter estimation analysis also makes a further method of subtraction validation possible for specific waveforms and instrument states. These injection waveforms are calculated from some set of known source parameters, which is used to test the parameter estimation analyses. In this way, we can test whether this noise subtraction scheme is legitimately reducing unwanted technical noise without distorting the measured gravitational-wave signals. We use the DeepClean algorithm to filter noise from a stretch of data which contains an astrophysical software injection. Then we check that the resultant posterior parameter estimation distributions are consistent with those from the pre-subtraction strain signal and not significantly biased away from the known injected parameters.

In this analysis, `DeepClean` co-author Tri Nguyen [37] injected non-spinning binary black hole (BBH) signals into the gravitational-wave strain and compared the recovered source parameters from the cleaned and uncleaned time-series.

The signals were injected beginning on Jun 25, 2017 07:39:48 UTC. Both BBH component masses were drawn from a uniform distribution $[28, 64]M_{\odot}$ and a mass ratio $q \in [0.125, 1.0]$ was imposed. This follows the analysis in [104]. The sky coordinates θ and ϕ, as well as the luminosity distance D_L were also sampled uniformly. The SNR of the resulting signals ranged from 1.5 to 18.7.

For the O3 60 Hz dataset, 12 high-mass BBH signals ($100 - 300M_{\odot}$) were injected across the $\sim$8 day period beginning on May 31, 2019 03:37:58 UTC. The signals were selected such that they contribute significant power near 60 Hz making the visualization and recovery more clear. The waveform models for the injections during O2 and O3 were generated from the waveform approximant `IMRPhenomPv2` [105, 106, 107].

We applied Bayesian statistics and estimated the posterior probability distribution of the source parameters using the Dynamic Nested Sampling algorithm `Dynesty` [108] implemented in the `Bilby` library [82]. The posterior distribution was estimated using only the gravitational-wave strain from a single detector, i.e. LIGO Hanford. The recovered quantities in this study are the component masses, inclination angle and luminosity distance.

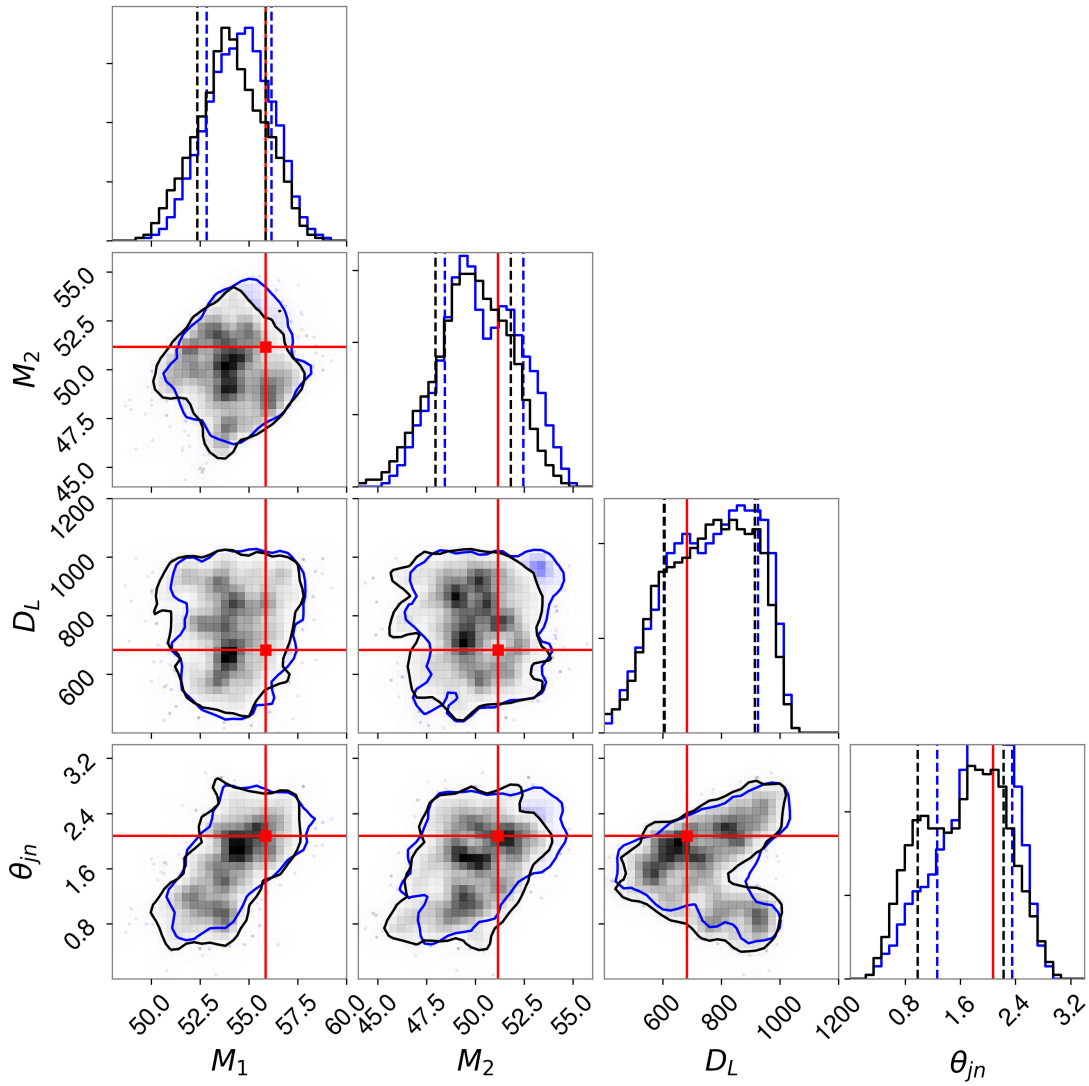

Figure 5.10: Corner plot showing the parameter estimation of the O2 data from the Hanford detector after cleaning the data with the DeepClean neural network and using the same auxiliary channel list as was used in the linear cleaning analyses. The recovered signals are injected via software and are not natural events. The true parameters of these injections are denoted by the red lines. The recovered data is given by the blue histograms and contours, whereas the unprocessed data is shown in black. The contours represent the 90% credible intervals and the vertical lines on the histograms represent a 1σ variation from the mean. As can be seen, all of the injected parameters are faithfully recovered. This image is used with permission from T. Nguyen [37]

.

We find that all of the quantities are reconstructed within 3σ of their true injected values. For the 'jitter' noise in the O2 dataset, it can be seen that the 90% credible intervals of the recovered quantities are decreased by $\sim$7% relative to the uncleaned recovery. In addition, we observed an average increase in the recovered optimal SNR of about 21.6%. Figure 5.10 shows a posterior distribution recovered from an example injection in the O2 linear dataset. There was not however any significant decrease observed in the 90% credible intervals or reconstructed SNR for the O3 60 Hz dataset. This is not surprising given that subtracting only the 60 Hz line and its sidebands should not significantly change the ASD. In all injections, the parameters recovered from the cleaned strain were consistent with the true values and those recovered from the original strain, suggesting that the network did not introduce any noise or corrupt the gravitational-wave signals.

5.8 Outlook for Neural Networks at LIGO

In addition to analytic methods, neural networks such as DeepClean have demonstrated the extended advantage of being able to determine linear, non-linear and non-stationary couplings into the detector output without previous knowledge of the physical mechanisms of the noise. The ability of the MLAs to successfully subtract non-linear couplings allows for network-derived filters to become more valuable as Advanced LIGO and future detectors become increasingly sensitive to additional, more complicated noise sources and the hardware engineering limit is approached. In addition, software upgrades are significantly less expensive in time and dollars relative to hardware commissioning, and therefore serve as the desired way of increasing the sensitivity and success of the experiment.

It is possible at this point to begin running DeepClean in real time, thereby facilitating use to multiple pipelines throughout the LIGO collaboration. If the networks are able to model and subtract the effects of certain nonlinear and non-stationary noise sources such as that of scattered light which was a significant issue during O3, they could become the first line of defence for low-latency and offline subtraction data quality efforts. The ability to capture these more complicated noise sources may be enough to promote marginal event candidates [2] to fully confident detections.

There is a notable downside to running deep neural nets for regression, particularly for "blind regression." There are $\mathcal{O}(10^4)$ auxiliary channels at each LIGO detector site. For a network which trains in a modest 5 minutes, finding bilinear noise by randomly choosing channels could take $\sim 10^3$ years to find. Clearly this is beyond the pale and smarter methods of channel selection must be made. The networks are an invaluable tool that successfully parameterize our ignorance of the detector noise and will eventually become the primary filter for the data, however the results are only as good as the data that can be provided. It is therefore important to continue to develop new channel detection and selection methods and to continue hardware injection and coupling tests. One such software method which will be discussed in chapter 6 is to develop lightweight linear, nonlinear, and non-stationary filters which are fast enough to brute force search through the channel parameter space in parallel and tell the networks which type of coupling it should be looking for.

Chapter 6

Towards Nonlinear and Non-stationary Filters

Gravitational wave detectors are highly sensitive machines and becoming increasingly more so with time due to developments in software techniques and hardware upgrades [109, 38, 13, 110]. As the noise floor continues to be pushed downwards, noise couplings are emerging which are not linear or stationary within te interval of investigation. Therefore, in order to sustain the rate of increase in the sensitivity of LIGO and future detectors, it is important to develop and implement methods which can filter out non-stationary and nonlinear noise from the data readout. In this chapter, we present methods for subtracting nonlinear and non-stationary noise couplings from the gravitational wave detector differential arm length readout measurement using the auxiliary monitors for training of the filters.

6.1 Wiener Filters

The predominant coupling order for noise sources into DARM is linear. Figure 2.2 showing the O3 noise budget plot is constructed by determining the linear coupling order of each noise term and producing a weighted sum of the individual noise curves. Above $\sim$100 Hz the estimation of the noise floor closely matches the measured curve. During aLIGO's second observing run (O2) however, this was not the case. The $\sim 100-1000$ Hz regime was primarily limited by beam jitter caused by a water cooling pump on the

beam table which was removed before O3. Fortunately, the jitter coupling was linear and well-observed by a number of auxiliary monitors (see figure 6.1) and therefore was able to be regressed. In this section, we will construct an optimal linear mean-square-error (MSE) filter known as a "Wiener filter" and apply it to the O2 data to remove the beam jitter noise. This simple filter will act as a starting point from which we will adapt and modify the signal processing techniques such that higher order effects and nonlinear source dynamics are effectively captured.

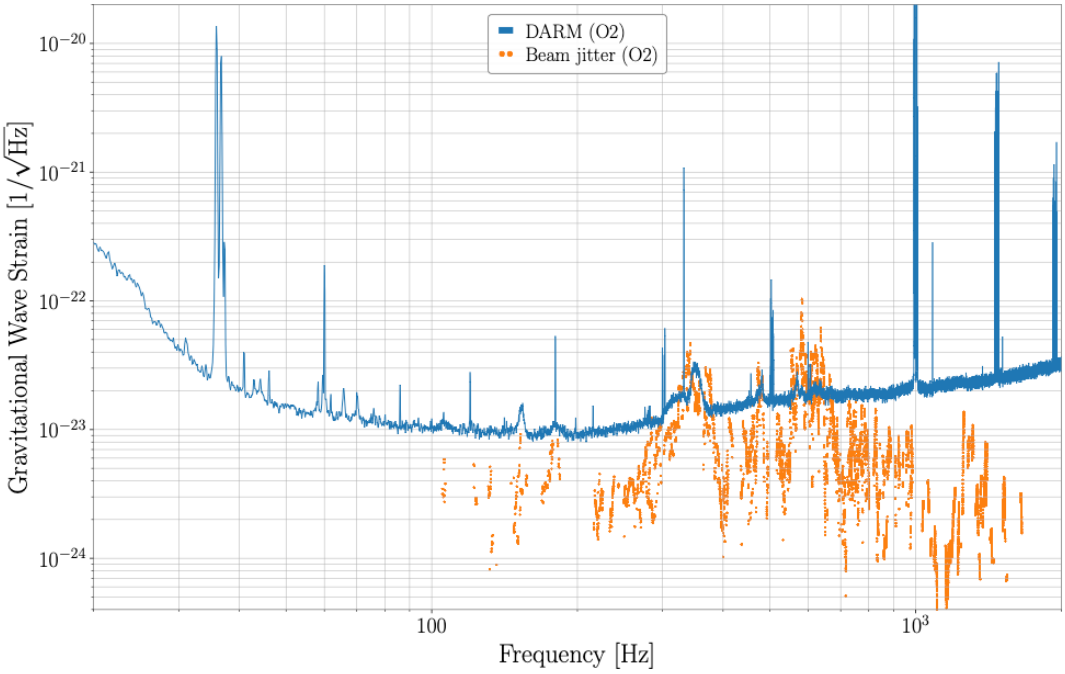

Figure 6.1: Broadband coupling of the beam jitter during aLIGO's second observing run. There is a noticeable bump in the spectrum from ∼300−400 Hz and some slightly narrower peaks near 500 Hz. While difficult to tell, the entire noise floor is actually raised which will become evident after regression of the beam jitter noise.

To begin, let us assert that we have the system signal $d[k]$ and we have an estimate of the noise, which we assume to be wide-sense stationary, to the system given by the convolution

$$y[k] = \sum_{m=0}^{M-1} a_k[m]x[k-m] \tag{6.1}$$

where $a_k[m]$ is the filter coefficient at the k^{th} time step and at tap delay m and the signal $x[k-m]$ is the witness channel to the noise in the system signal at time step k and delay tap delay m. The tapped delay line allows for a lag to exist between the data streams and increases the number of weights used to characterize the noise. Since the witness channel used is known, once the weights are known, they can be convolved with the witness channel signal using equation (6.1) to produce an estimate of the noise due to the witness channel. This estimate, $y[k]$, can then be subtracted away from the system signal $d[k]$.

In order to find the appropriate filter coefficients, we need to minimize the error signal with respect to the coefficient filter taps according to some cost function metric. In this case, we seek to minimize the expectation value of the the MSE. This can be expressed by the following

$$\begin{aligned}
\frac{\partial}{\partial a_k[n]}\big\langle d[k]-y[k]\big\rangle^2 &= \frac{\partial}{\partial a_k[n]}\big\langle -2d[k]y[k]+y^2[k]\big\rangle \\
&= \Big\langle -d[k]x[k-n]\Big\rangle + \left\langle \sum_{m=0}^{M-1} a_k[m]x[k-m]x[k-n]\right\rangle \\
&= 0
\end{aligned} \tag{6.2}$$

The first term of the RHS of the second line is the cross-correlation vector of the witness channel with the output strain with length M. The second term on this line is the weighted $M \times M$ auto-correlation matrix of the witness channel. We may represent this matrix equation as

$$\begin{pmatrix} p[0] \\ p[1] \\ \vdots \\ p[M-1] \end{pmatrix} = \begin{pmatrix} R[0] & R[1] & \cdots & R[M-1] \\ R[1] & R[0] & \cdots & R[M-2] \\ \vdots & & \ddots & \vdots \\ R[M-1] & R[M-2] & \cdots & R[0] \end{pmatrix} \begin{pmatrix} a[0] \\ a[1] \\ \vdots \\ a[M-1] \end{pmatrix} \tag{6.3}$$

Equivalently, if we define

$$p[n] \equiv \langle d[k]x[k-n] \rangle$$
$$R[m-n] \equiv x[k-m]x[k-n] \tag{6.4}$$

then we may write equation (6.3) as

$$p[k] = \sum_{m=0}^{M-1} a_k[m]R[m-n] \tag{6.5}$$

Or in vector notation this becomes

$$\mathbf{p} = \hat{\mathbf{R}}\mathbf{a} \tag{6.6}$$

Therefore, we can simply find the optimal coefficients for the filter by inverting the auto-correlation matrix and multiplying it by the cross-correlation vector giving

$$\mathbf{a} = \hat{\mathbf{R}}^{-1}\mathbf{p} \tag{6.7}$$

The filter construction just described convolves the weights of a single input channel to produce a single output channel. This is known as a "single-in single-out" (SISO) filter. For uncorrelated auxiliary channels, this filter is sufficient. If multiple noise sources couple independently and linearly into the output data, the SISO filter may be run over each channel to produce a set of noise estimates which can then be subtracted from the output signal. The restriction to uncorrelated input witness channels turns out to be a strict one. Consider a 3-axis seismometer: this sensor will return three data streams - one from each direction. These three data streams will, in general, be correlated. If this is the case, then the optimal filter cannot be found by inspecting a single channel direction at a time. Therefore, we seek to find a filter construction which can optimize the weight coefficients for a set of input channels at the same time. In other words, we would like the filter to be "multiple-in single-out" (MISO). To begin, let us extend equation (6.1) to account for multiple input channels

$$y[k] = \sum_{b=0}^{B-1} \sum_{m=0}^{M-1} a_k^{(b)}[m] x^{(b)}[k-m] \tag{6.8}$$

where we are considering B channels with a delay line of M taps. Following equation (6.2) to minimize the mean square error, $e[k]$, we find

$$\frac{\partial \left\langle e^2[k] \right\rangle}{\partial a_k^{(b')}[j]} = -2 \left\langle d[k] \begin{pmatrix} x^0[k-j] \\ x^1[k-j] \\ \vdots \\ x^{B-1}[k-j] \end{pmatrix} \right\rangle + 2 \left\langle y[k-j] \begin{pmatrix} x^0[k-j] \\ x^1[k-j] \\ \vdots \\ x^{B-1}[k-j] \end{pmatrix} \right\rangle \tag{6.9}$$

which can be written more explicitly as

$$\begin{aligned} \frac{\partial \left\langle e^2[k] \right\rangle}{\partial a_k^{(b')}[j]} &= -2 \left(p^{(0)}[j] \delta^{b',0} + p^{(1)}[j] \delta^{b',1} + \cdots \right) \\ &= \left\{ \sum_{m=0}^{M-1} a_k^{(0)}[m] R^{(0,0)}[m-j] \delta^{b',0} + \sum_{m=0}^{M-1} a_k^{(1)}[m] R^{(1,0)}[m-j] \delta^{b',0} + \cdots \right. \\ &= \sum_{m=0}^{M-1} a_k^{(0)}[m] R^{(0,1)}[m-j] \delta^{b',1} + \sum_{m=0}^{M-1} a_k^{(1)}[m] R^{(1,0)}[m-j] \delta^{b',1} + \cdots \\ &= + \cdots \\ &= \left. + \sum_{m=0}^{M-1} a_k^{(0)}[m] R^{(0,B-1)}[m-j] \delta^{b',B-1} + \sum_{m=0}^{M-1} a_k^{(1)}[m] R^{(0,B-1)}[m-j] \delta^{b',B-1} + \cdots \right\} \end{aligned} \tag{6.10}$$

where δ^{ij} is a Kronecker delta that arises as a result of taking the derivative of the weight kernel $a_k^{(b)}[m]$. Here we can see the effect of including multiple input channels at once. Namely, the terms with $R^{(q,s)}[m-j]$ where $q \neq s$ are now cross-correlation matrices which were assumed in the SISO case to vanish. Because these terms multiply the filter coefficients, the filter is adjusted due to these non-vanishing cross-correlations. We may write equation (6.10) in matrix form giving

$$\begin{pmatrix} \mathbf{p}^{(0)} \\ \mathbf{p}^{(0)} \\ \vdots \\ \mathbf{p}^{(B-1)} \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{R}}^{(0,0)} & \hat{\mathbf{R}}^{(1,0)} & \dots & \hat{\mathbf{R}}^{(B-1,0)} \\ \hat{\mathbf{R}}^{(0,1)} & \hat{\mathbf{R}}^{(1,1)} & \dots & \hat{\mathbf{R}}^{(B-1,1)} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathbf{R}}^{(0,B-1)} & \hat{\mathbf{R}}^{(1,B-1)} & \dots & \hat{\mathbf{R}}^{(B-1,B-1)} \end{pmatrix} \begin{pmatrix} \mathbf{a}^{(0)} \\ \mathbf{a}^{(1)} \\ \vdots \\ \mathbf{a}^{(B-1)} \end{pmatrix} \tag{6.11}$$

where each $\mathbf{p}$ and $\mathbf{a}$ is a vector and each $\hat{\mathbf{R}}$ is an $M \times M$ matrix. Therefore the complete cross-correlation and filter coefficient 'vectors' have dimensions $B \times M$ and the full correlation matrix has dimensions $B \times B$ where each entry $\hat{\mathbf{R}}$ is an $M \times M$ matrix. If each off-diagonal cross-correlation matrix were to vanish, we would have a set of B separable equations that corresponds to B iterations of the SISO Wiener filter. We can therefore construct the optimal MSE filter using correlated input channels. The python code to implement this extended Wiener filter is given is appendix A.1.

To continue the example, for the O2 linear regression 23 channels were used [103]. They include:

Pre-Stabilized Laser

- `PSL-DIAG_BULLSEYE_PIT_OUT_DQ`: "Bullseye" phodiode pitch
- `PSL-DIAG_BULLSEYE_YAW_OUT_DQ`: "Bullseye" photodiode yaw
- `PSL-DIAG_BULLSEYE_WID_OUT_DQ`: "Bullseye" photodiode roll

Alignment Sensing and Control

- `ASC-DHARD_P_OUT_DQ`: "D-hard" mode pitch
- `ASC-DHARD_Y_OUT_DQ`: "D-hard" mode yaw
- `ASC-CHARD_P_OUT_DQ`: "C-hard" mode pitch
- `ASC-CHARD_Y_OUT_DQ`: "C-hard" mode yaw

Input Mode Cleaner

- `IMC-WFS_A_DC_PIT_OUT_DQ`: Wavefront sensor "A" pitch
- `IMC-WFS_A_DC_YAW_OUT_DQ`: Wavefront sensor "A" yaw
- `IMC-WFS_B_DC_PIT_OUT_DQ`: Wavefront sensor "B" pitch
- `IMC-WFS_B_DC_YAW_OUT_DQ`: Wavefront sensor "B" yaw

Length Sensing and Control

- `LSC-PRCL_IN1_DQ`: Power recycling cavity length
- `LSC-SRCL_IN1_DQ`: Signal recycling cavity length
- `LSC-MICH_IN1_DQ`: Michelson arm cavity length
- `LSC-CAL_LINE_SUM_DQ`: Calibration line modulation

Physical Environment Monitor

- `PEM-EY_MAINSMON_EBAY_1_DQ`: Voltage monitor in y-arm electronics bay 1
- `PEM-EY_MAINSMON_EBAY_2_DQ`: Voltage monitor in y-arm electronics bay 2
- `PEM-EY_MAINSMON_EBAY_3_DQ`: Voltage monitor in y-arm electronics bay 3

Calibration Monitor

- `CAL-PCALY_EXC_SUM_DQ`: Y-arm excitation
- `CAL-PCALX_EXC_SUM_DQ`: X-arm excitation
- `CAL-PCALY_TX_PD_OUT_DQ`: Y-arm $\theta(x)$ photodiode reading
- `CAL-CS_LINE_SUM_DQ`: Corner station calibration sum

The `PEM` and `CAL` channels are used to remove the 60 Hz electrical lines and self-inflicted calibration lines respectively. The remainder of the channels contribute to the removal of the broadband jitter noise, most importantly the `PSL` and `ASC` channels. Running this list of channels through equation (6.8) and subtracting the noise estimate from the strain channels produces the 'DARM (O2) After WF' curve in figure 6.2. Using the BNS range calculation from equation (2.2), it can be shown that the detector range increase from Wiener filtering is $20-50\%$ contingent upon the stretch of data being analyzed.

While the MISO WF outperforms the SISO WF, the effects are not particularly dramatic as the correlation between the auxiliary channels chosen for the O2 cleaning efforts is generally quite low. To fully exploit the benefits of the MISO filter, let us consider the example referenced above where seismic noise is the dominant contribution. We can create a simple mock signal $d(t) = s(t) + r(t) + n_i(t)$ where $s(t)$ is a 22 Hz sinusoid representing the target signal, $r(t)$ is white noise and $n_i(t)$ is correlated seismic noise from the $\hat{x}$, $\hat{y}$ and $\hat{z}$ channels of a single seismometer. The seismic channels are characterized in figure 6.3 and the results of the MISO vs. SISO filters is shown in

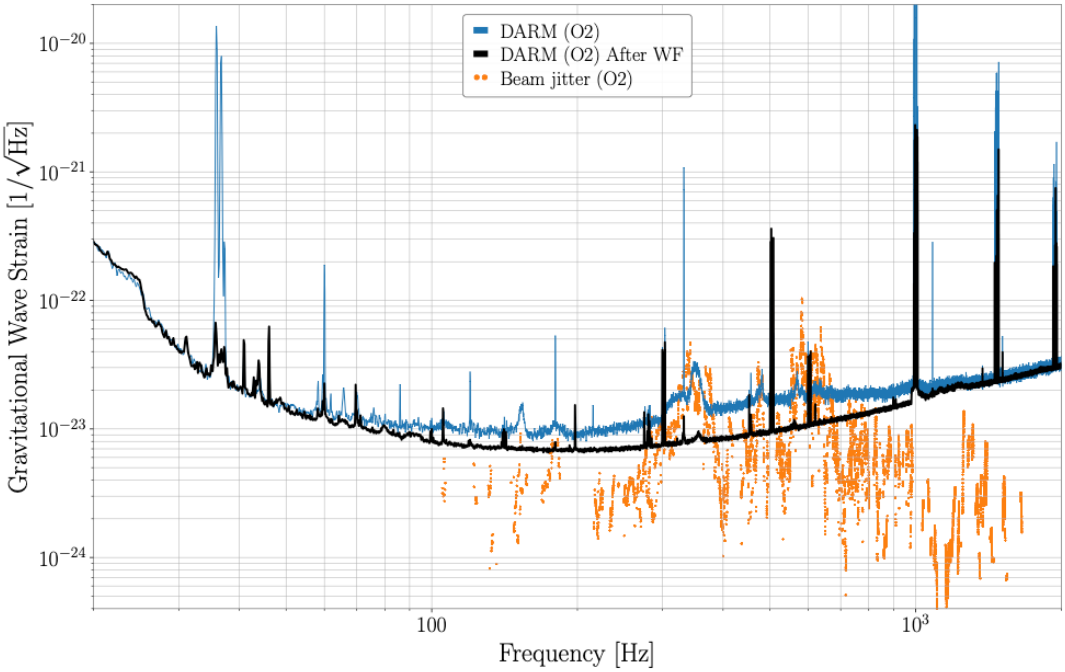

Figure 6.2: Subtraction of the broadband jitter noise using the generalized Wiener filter. The blue curve is the PSD of the strain data before running the filter and the black curve is the PSD after the subtraction has been performed in the time domain. In addition to the broadband subtraction, the removal of the calibration lines near 36 Hz and that of the power lines 60 Hz harmonics are also visible.

figure 6.4. While results on real data are generally less dramatic due to the wide array of noise sources contributing non-negligibly to the target signal, this example demonstrates the value in retaining the full correlation matrix. Similar results on correlated magnetic and seismic noise may be found in [111, 112].

6.2 Linear Adaptive Filters

The success of the MISO Wiener filter naturally leads to variants and extensions of this algorithm. One of the drawbacks of the Wiener filter is that is assumes that the noise is stationary during the period of training and testing. As we have seen in chapter 2 however, this is not in general true. The LIGO noise background changes randomly with the environment and detector status and therefore the noise contributions to the strain output are, in part, non-stationary. Here we consider non-stationary noise sources which last $\gtrsim 1$ minute. These noise sources whose frequency is a function of time are referred to as "wandering lines" to reflect their appearance is spectrograms.

6.2.1 Short Duration Iterative Wiener Filter

For a linear noise source with non-stationarity on the order of $t-$seconds, it may be assumed that for times $\ll t$, the noise is effectively stationary. Therefore we can propose to regress the noise iteratively for time intervals $\ll t$ with the SISO Wiener filter. There is however an immediate down-side to this approach - the performance of the filter may vary from interval to interval causing discontinuities in the time domain data which translate to spectral artifacts in the frequency domain. The result of the filter can therefore *add* noise to the system. Although it is possible to rescue this approach, it will not be considered further here. However, if the filter coefficients were updated at *every* time step as opposed to splitting the data into longer segments, this continuity issue would too vanish.

6.2.2 Least Mean Squared

Successfully characterizing non-stationary noise requires filter updates on timescales shorter than the duration of the non-stationarity. This aspect can be addressed by performing instantaneous updates at every time step. While any loss function can be

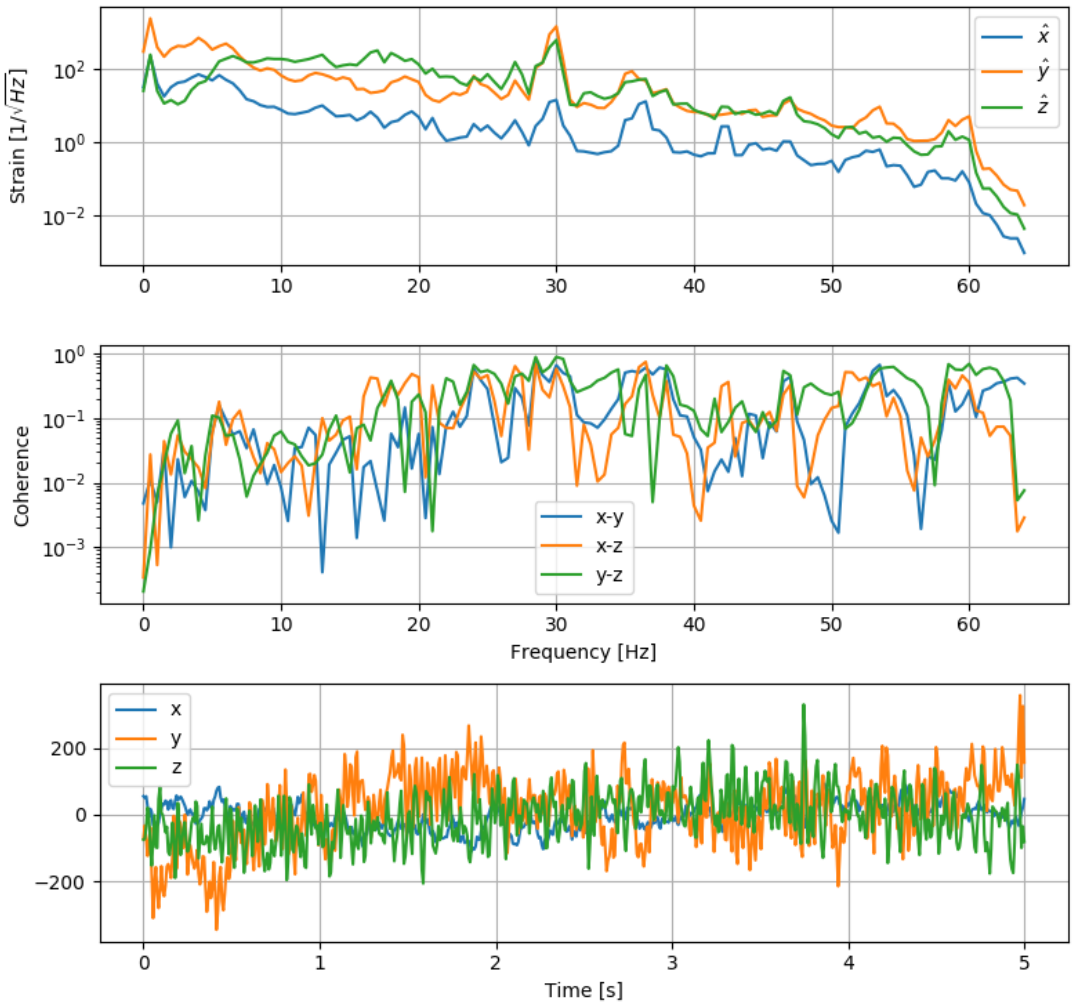

Figure 6.3: Comparison of the channels `H1:PEM-CS_SEIS_LVEA_VERTEX_X_DQ`, `H1:PEM-CS_SEIS_LVEA_VERTEX_Y_DQ` and `H1:PEM-CS_SEIS_LVEA_VERTEX_Z_DQ` from a single seismometer recorded during LIGO's second observing run. The top panel shows the PSD of the channels, the middle panel shows the coherence between the channels and the bottom panel is the mean-subtracted time series that is fed into the filters. Is is evident that there is correlation between the different channel directions in the seismometer which suggests usage of a MISO filter over looping the SISO filter over individual arrays to increase the data quality.

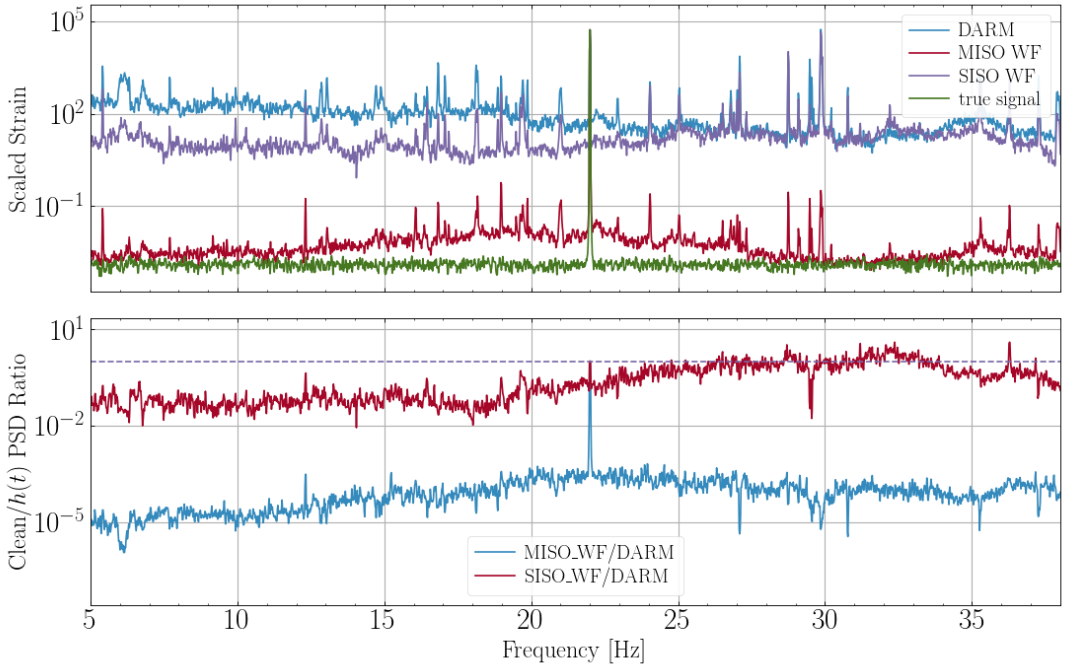

Figure 6.4: MISO versus SISO performance on correlated seismic noise. Due to the high degree of correlation between the channels, i.e., the off-diagonal elements of $\hat{\mathbf{R}}$ in equation (6.11) are comparable to the diagonal elements, the performance of the MISO filter is markedly improved over the SISO filter. The unobserved 22 Hz target signal and white noise have not been degraded by either filter.

used, here the least-mean-squared (LMS) loss function is used. If we wish do minimize the error signal, then we need to update the weights of the current step and apply them to the next step in such a way that the weights follow the negative gradient of the error signal. That is, given a weight vector $\mathbf{a}[k]$ of $M+1$ taps long, and a step size μ, we wish to calculate

$$\mathbf{a}[k+1] = \mathbf{a}[k] - \mu\vec{\nabla}e^2[k] \tag{6.12}$$

where the error signal is the difference of the target signal and noise estimate, $e[k] = d[k] - y[k]$. Filters that update instantaneously in this way are able to change to characterize the fluctuations of the witness signal and are referred to as *adaptive* filters. For a linear adaptive filter, we may use the noise estimate given by equation (6.1) to calculate the weight update equation. Feeding in equation (6.12) we find

$$\frac{\partial e^2[k]}{\partial a_k[n]} = 2e[k]\frac{\partial e[k]}{\partial a_k[n]} = -2e[k]\frac{\partial y[k]}{\partial a_k[n]} = -2e[k]x[k-n] \tag{6.13}$$

where $x[k]$ is the input noise signal vector given as

$$\mathbf{x}[k] = \begin{pmatrix} x[k] \\ x[k-1] \\ \vdots \\ x[k-M+1] \end{pmatrix} \tag{6.14}$$

Thus we find that the filter coefficients update as

$$\mathbf{a}[k+1] = \mathbf{a}[k] + 2\mu e[k]\mathbf{x}[k] \tag{6.15}$$

For data with a roughly flat noise power spectrum, equation (6.15) is sufficient. However, for a varying spectrum, only the frequencies with the most power are likely to be updated as they will dominate the error signal. If the spectral features we wish to address contain less power relative to the surrounding frequencies, we will have to normalize the spectrum. This can be done by whitening the data, however we must also de-whiten the data which poses a problem due to the zeros in the whitening window. Instead, we can locally normalize the data just in the $M+1$ taps by dividing the

error signal with the power in the witness signal. In other words, we prote the RHS of equation (6.15) by

$$2\mu e[k]\mathbf{x}[k] \longrightarrow \frac{2\mu e[k]\mathbf{x}[k]}{\mathbf{x}^T[k]\mathbf{x}[k]} \tag{6.16}$$

As a final modification, we consider the effect of very short, high power glitches in the witness channel.

It is not uncommon for auxiliary channels to glitch or be bumped or subjected to some transient, stochastic influence causing a misreading of the true environmental data. In figure 6.5 we created a mock data set where the witness channel has a sine-gaussian glitch at 1.6 seconds and a D.C offset from 2.5−3 seconds. The witness channel is clearly not a good estimate of the system signal during these glitches and offsets. The estimate from the LMS filter design given by equation (6.15) fails during these periods and eventually the weights recover and the subtraction is again adequate. To alleviate the effect of glitching, we can restrict the amount of the current gradient update that passes forward to the next time step. Specifically, we can write

$$\mathbf{a}[k+1] = (1-\alpha)\mathbf{a}[k] - \frac{2\mu}{\mathbf{x}^T[k]\mathbf{x}[k] + \psi} e[k]\mathbf{x}[k] \tag{6.17}$$

where α is the "leak" parameter which determines how much of the weights to pass forward, and ψ is a safety term to prevent division errors. The code is available in appendix A.2.

This leaky, normalized LMS filter (LNLMS) has direct applications for LIGO data. For the first six months of O3, frequency dependent "wandering" lines appeared in the $\sim 80 - 110$ Hz band and at integer multiple harmonics of DARM as seen in figure 6.6. It was found that noise from the light squeezers was coupling into DARM. Fortunately, these subsystems are well monitored and this noise can be regressed using a linear adaptive filter. Focusing on the fundamental harmonic and applying equation (6.17) to the squeezer noise results in figure 6.7.

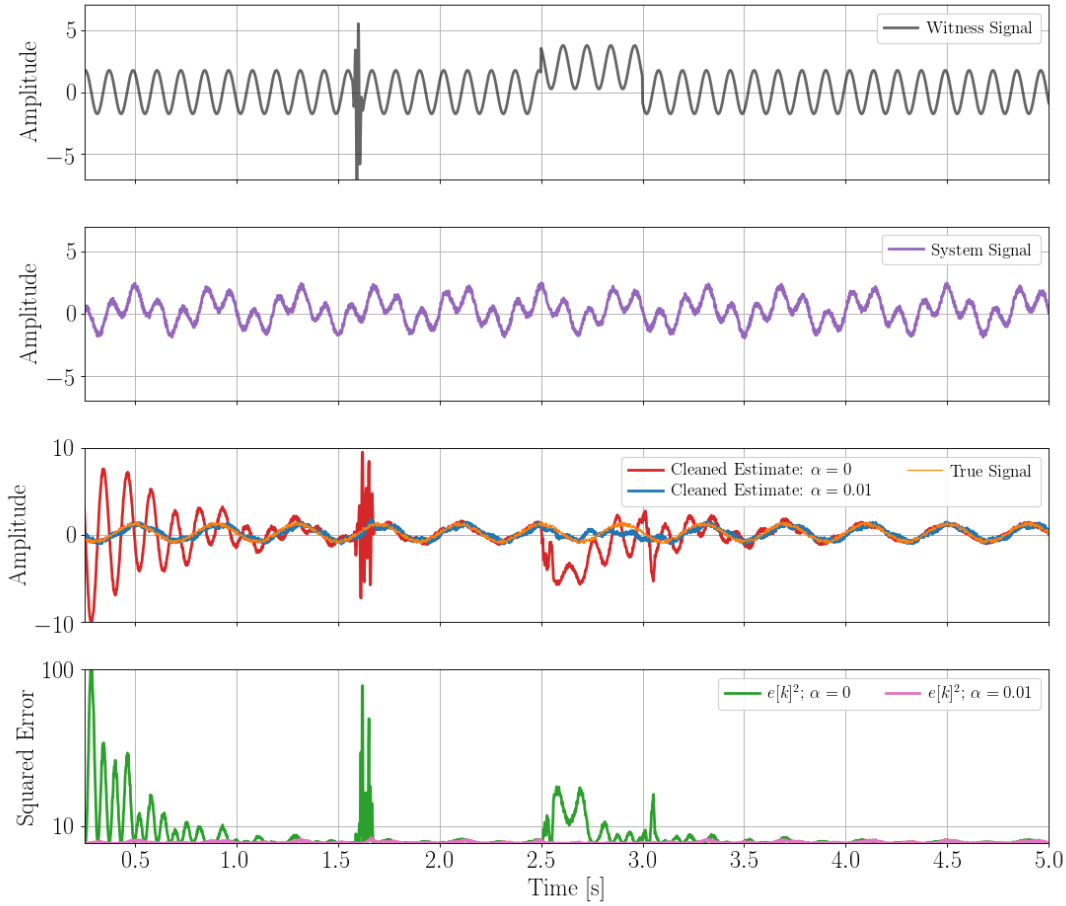

Figure 6.5: Performance test of the leak parameter α for a normalized least mean squares filter. The witness signal (top plot) has a high frequency and high amplitude sine-gaussian glitch at 1.6 s and a D.C offset from 2.5−3.0 s. Regressing the system signal with the witness channel adds noise to the system output starting at the time of the glitch and lasting until the weights are retrained after enough iterations. However, using a non-zero leak parameter has the effect of dampening transients introduced into the witness channel. For highly non-stationary noise this must be adjusted to compensate for the changing frequency dependence.

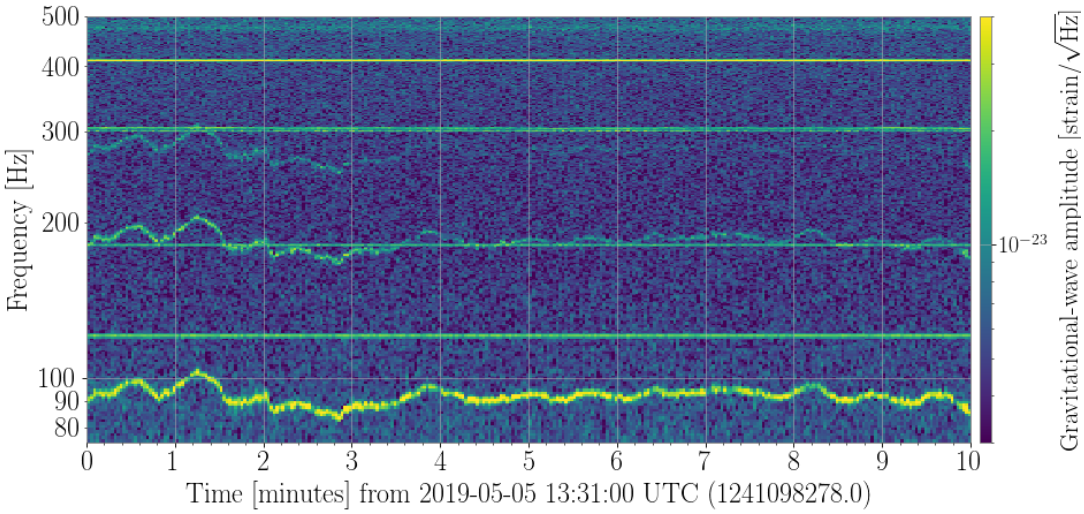

Figure 6.6: The laser's relaxation oscillation is suppressed by a built-in power stabilization mechanism known as the "noise eater" and is used in conjunction with photon squeezing techniques [113, 71]. During O3, this noise eater created time dependent frequency responses that coupled into DARM near 100 Hz and resonated at integer harmonics up the spectrum.

6.3 Nonlinear Filters

In addition to the linear stationary and non-stationary noise sources, it is expected, and is indeed the case, that nonlinear noise manifests in the strain data. Because of the difficulty in diagnosing nonlinear noise, it is not clear exactly how many nonlinear coupling are transient or persistent in the data, or the order of nonlinearity of the couplings. Some sources are able to be identified however. The removal of such terms necessitates a filter with the required order of nonlinearity built in. In this section, we shall develop bilinear filters, also called Volterra filters, and show how they can be extended to higher orders.

First, we can perform an expansion of the filter estimate in powers of the auxiliary channels. To second order, this is

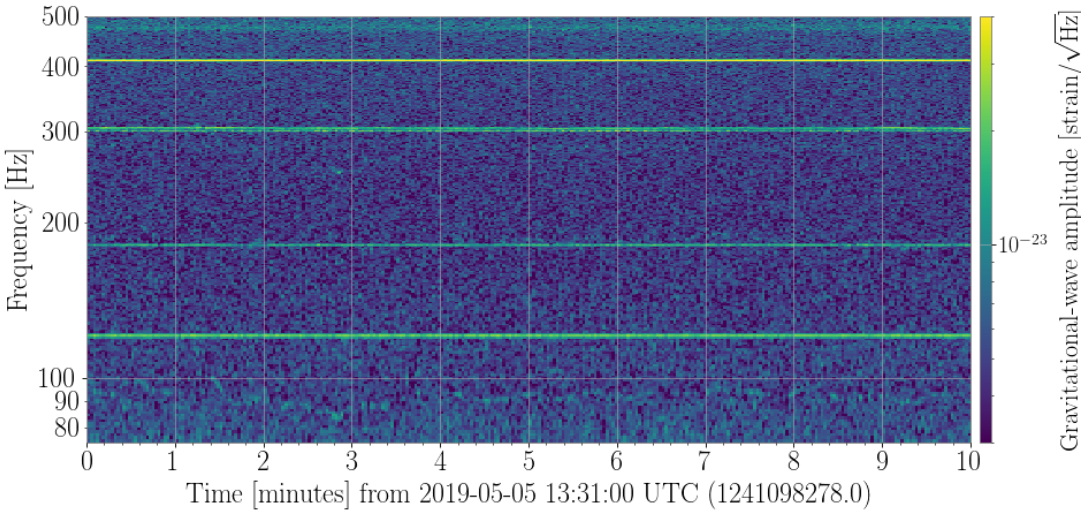

Figure 6.7: Spectrogram of DARM after filtering demonstrating the efficiency of the LNLMS algorithm at removing the time dependent frequencies. Comparing with figure 6.6, we see that the fundamental signal as well as the higher order harmonics have been removed.

$$y[k] = b_0 + \sum_{m=0}^{M-1} a_k[m]x[k-m] + \sum_{m=0}^{M-1}\sum_{m'=0}^{M-1} a_k[m,m']x_1[k-m]x_2[k-m'] \tag{6.18}$$

The first term on the RHS is a DC offset, the second term is the linear expansion used to create the Wiener filter and the third term is bilinear and the one for which we will develop a bilinear filter.

6.3.1 Volterra Filter

A generalized second order filter may be constructed assuming a bilinear noise estimate given by

$$y[k] = \sum_{m=0}^{M-1}\sum_{m'=0}^{M-1} a_k[m,m']x_1[k-m]x_2[k-m'] \tag{6.19}$$

We can shadow the procedure for the Wiener filter and optimize the filter coefficients using the MSE cost function. To that end, we take the derivative of the MSE with respect to the filter coefficients giving

$$\left\langle d[k]x_1[k-i]x_2[k-j] \right\rangle = \sum_{m=0}^{M-1}\sum_{m'=0}^{M-1} a_k[m,m'] \left\langle x_1[k-m]x_2[k-m']x_1[k-i]x_2[k-j] \right\rangle \tag{6.20}$$

The term on the LHS is just a three-point cross-correlation matrix of the input signals multiplied by the system signal. The term on the RHS is a four-point correlation of the input channels which cannot be split into products of two-point correlations because LIGO data is in general not stationary or Gaussian. Adopting notation where

$$\begin{aligned} C^{(i,j)}[m,m'] &\equiv \left\langle x_1[k-m]x_2[k-m']x_1[k-i]x_2[k-j] \right\rangle \\ P^{(1,2)}[i-j] &\equiv \left\langle d[k]x_1[k-i]x_2[k-j] \right\rangle \end{aligned} \tag{6.21}$$

allows equation (6.20) to be written as

$$P^{(1,2)}[i-j] = \sum_{m=0}^{M-1}\sum_{m'=0}^{M-1} a_k[m,m']C^{(i,j)}[m,m'] \tag{6.22}$$

The LHS is a matrix of size $M \times M$ and therefore the RHS must be as well. Look however at $C^{(i,j)}[m,m']$. For each (i,j) pair that we pick, we are still left with an $M \times M$ matrix from the m and m' terms. The filter coefficients are also a matrix of size $M \times M$. In order to get a scalar then, it seems like we need to "dot" these matrices. This is hard to understand, though the direct product is easy to code (albeit expensive to run). There is another way to visualize the algebra here; instead of a matrix of matrices dotting yet another matrix, we flatten the $C^{(i,j)}[m,m']$ matrix to a vector of size M^2 and we do the same to the coefficients. For example, if i,j,m and m' all run from 0 to 1, then

$$\hat{\mathbf{a}} = \begin{pmatrix} a[0,0] & a[0,1] \\ a[1,0] & a[1,1] \end{pmatrix} \rightarrow \begin{pmatrix} a[0,0] \\ a[0,1] \\ a[1,0] \\ a[1,1] \end{pmatrix} \tag{6.23}$$

Looking at equation (6.23) we can start running through i and j and do the sum. The $i = 0 = j$ term would be

$$\begin{aligned} P^{(1,2)}[0] = C^{(0,0)}[0,0]a[0,0] + C^{(0,0)}[0,1]a[0,1] \\ + C^{(0,0)}[1,0]a[1,0] + C^{(0,0)}[1,1]a[1,1] \end{aligned} \tag{6.24}$$

Similarly, the $i = 1$, $j = 0$ term would be

$$\begin{aligned} P^{(1,2)}[1] = C^{(1,0)}[0,0]a[0,0] + C^{(1,0)}[0,1]a[0,1] \\ + C^{(1,0)}[1,0]a[1,0] + C^{(1,0)}[1,1]a[1,1] \end{aligned} \tag{6.25}$$

Collecting everything we see that we can write this as a matrix multiplication problem

$$\begin{pmatrix} P^{(1,2)}[0] \\ P^{(1,2)}[1] \\ P^{(1,2)}[1] \\ P^{(1,2)}[0] \end{pmatrix} = \begin{pmatrix} C^{(0,0)}[0,0] & C^{(0,0)}[0,1] & C^{(0,0)}[1,0] & C^{(0,0)}[1,1] \\ C^{(0,1)}[0,0] & C^{(0,1)}[0,1] & C^{(0,1)}[1,0] & C^{(0,1)}[1,1] \\ C^{(1,0)}[0,0] & C^{(1,0)}[0,1] & C^{(1,0)}[1,0] & C^{(1,0)}[1,1] \\ C^{(1,1)}[0,0] & C^{(1,1)}[0,1] & C^{(1,1)}[1,0] & C^{(1,1)}[1,1] \end{pmatrix} \begin{pmatrix} a[0,0] \\ a[0,1] \\ a[1,0] \\ a[1,1] \end{pmatrix} \tag{6.26}$$

Lastly, we can put this all in matrix notation to get

$$\begin{aligned} \vec{\mathbf{P}} &= \hat{\mathbf{C}}\vec{\mathbf{a}} \\ \rightarrow \vec{\mathbf{a}} &= \hat{\mathbf{C}}^{-1}\vec{\mathbf{P}} \end{aligned} \tag{6.27}$$

Since each matrix element of $\hat{\mathbf{C}}$ is composed of six terms, then there are a total of

$(4 \times 4) \times 6 = 96$ terms in the 2-tap case. More generally, there will be $(M^2 \times M^2) \times 6 = 6M^4$ terms for an $M-$tap filter. We can see an immediate limitation to this bilinear filter, the size of the matrix grows with the tap length as M^4 as opposed to the Wiener filter which grows as M^2. A tap length of 32 would lead to $\sim 10^6$ matrix elements which is about as big as we can reasonably invert in a usable amount of time. The python code for this filter is described in appendix A.3.

Armed with a method for calculating the bilinear filter coefficients, we may now apply it to real data which has been downsampled to 256 Hz. During O3, it was found that the 60 Hz power mains were beating against a 0.25 Hz angular mode ("DHARD") of the mirrors measured by the Angular Sensing and Control (ASC) system. This produced beat notes at $60 \pm 0.25n$ Hz where $n = 1, 2, 3 \ldots$. There is no single witness which observes these beat frequencies and therefore both channels must be used together to filter the noise. Running a two-tap Volterra filter in series with a Wiener filter over the data using one channels which measures the power mains and another which measures the slow DHARD modes of the mirrors, we find the result in figure 6.8; the bilinear noise is effectively removed without adding artifacts or noise back into the system. It can be observed that the 60 Hz line is not itself removed by the bilinear filter. This is because the Volterra filter is strictly second order and cannot remove linear artifacts.

6.3.2 Slow Bilinear Coupling

If the bilinear coupling of channels evolves slowly, that is, if the modulation of noise is slow relative to the frequency of the noise, then a simplification may be made to equation (6.19). Let us assume that the modulating frequency of $x_2[k]$ is much less than the relevant frequencies of $x_1[k]$. If this is true, then across a window M time steps wide (an M-tap filter), the amplitude of $x_2[k]$ is effectively unchanged. For example, if the modulating frequency of $x_2[k]$ is 0.5 Hz and it couples into $x_1[k]$ which oscillates at 60 Hz, then at a sample rate of 256 Hz for every 60 Hz cycle ($\sim$4 time steps), we see $4/256 \times 0.5 \times 100 < 1\%$ of the cycle of the modulating frequency. Under these conditions, we may treat the modulating frequency as stationary and write

$$y[k] \approx x_2[k] \sum_{m=0}^{M-1} a_k[m] x_1[k-m] \tag{6.28}$$

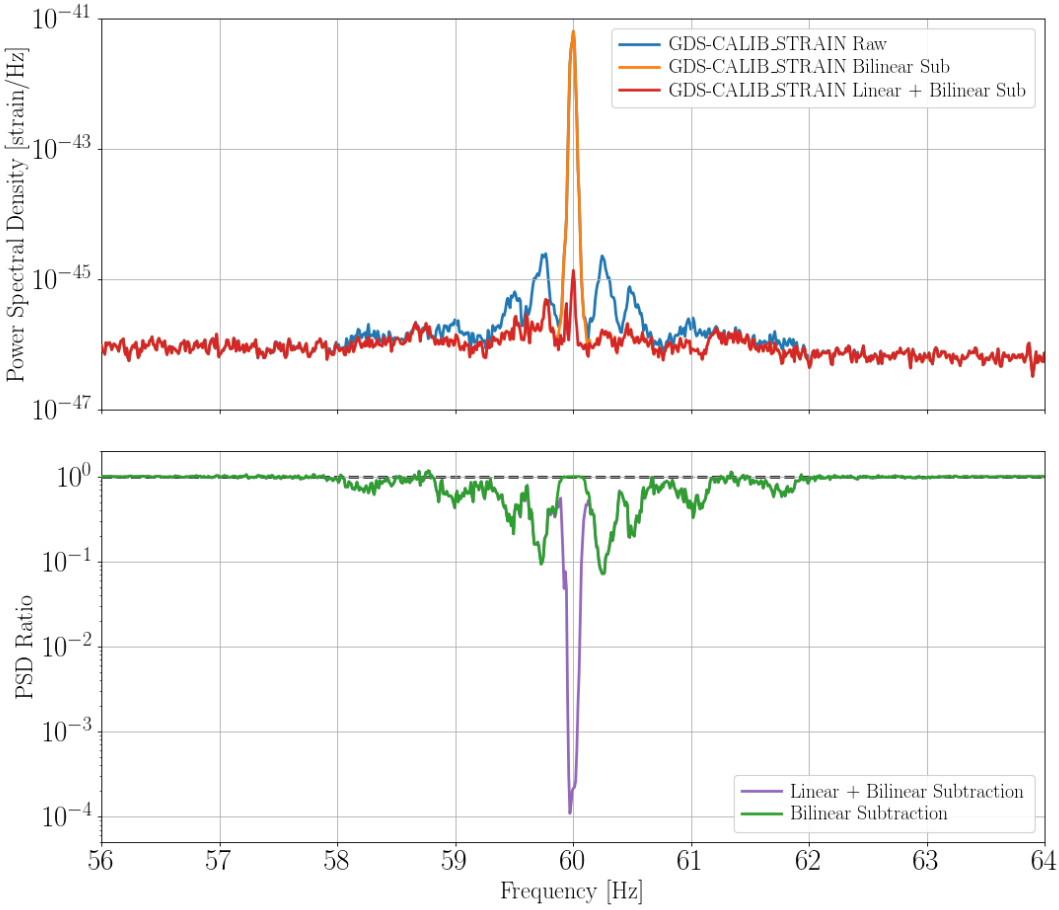

Figure 6.8: *Top:* Power spectral density of DARM at the Hanford site before and after the removal of the sidebands around the 60 Hz power lines. The orange curve is a result of applying a two-tap Volterra filter. The red curve was obtained by using a linear and bilinear filter in series. *Bottom:* Ratio of the PSD of the cleaned spectrum to the original spectrum shows the power that has been removed from the spectrum due to the beat notes. The green curve is the ratio of the bilinearly cleaned PSD to the raw PSD. The purple curve is the ratio of the linearly and bilinearly cleaned PSD to the raw PSD.

This reduces the Volterra filter to a modulated linear filter. Using instantaneous updates to create an adaptive filter capable of removing non-stationary noise gives

$$\mathbf{a}[k+1] = \mathbf{a}[k] + 2\mu e[k] x_2[k] \mathbf{x_1}[k] \tag{6.29}$$

A modulated linear adaptive filter of this kind is well suited to filter the sidebands caused by the beating of the DHARD modes of the mirror with the electrical lines as mentioned in section 6.3 and has been applied successfully[99].

6.4 Increasing Filter Order

As the noise floor is lowered due to hardware upgrades in gravitational wave detectors, higher order noise couplings are appearing. Developing $\mathcal{O}(x^2)$ analytic filters quickly becomes unwieldy and computationally expensive. While it is possible to search for arbitrarily nonlinear and non-stationary couplings with deep learning algorithms as in chapter 5, these are also computationally costly and time intensive. It would therefore be advantageous to develop lightweight analytical filters of arbitrary order which would be continually run in parallel across machines to find and remove these couplings.

6.4.1 Adaptive Volterra Filter and Beyond

We may make a second order adaptive filter by calculating the instantaneous update equation for equation (6.19). This is simply done and it can be found to be

$$a_{k+1}[m, m'] = a_k[m, m'] + 2\mu e[k] x_1[k-m] x_2[k-m'] \tag{6.30}$$

This may be cast into a matrix form as

$$\begin{aligned} \hat{\mathbf{a}}_{k+1} &= \hat{\mathbf{a}}_k + 2\mu e[k] (\mathbf{x_1}[k] \otimes \mathbf{x_2}^T[k]) \\ &= \hat{\mathbf{a}}_k + 2\mu e[k] \begin{pmatrix} x_1[k] \\ x_1[k-1] \\ \vdots \\ x_1[k-M+1] \end{pmatrix} \begin{pmatrix} x_2[k] & x_2[k-1] & \cdots & x_2[k-M+1] \end{pmatrix} \end{aligned} \tag{6.31}$$

where hats denote matrices. It is clear then that additional input channels may be

added and the weight update will depend on the outer product of the M taps of the input channels. This is however not as lightweight as we would like. Instead, one may begin with the linear LMS filter and choose to "warp" the input space by feeding in nonlinear data. In other words, the nonlinear effects of a time series may be captured by constructing the appropriate nonlinear input signal and using a linear filter in contrast to using linear input signals and constructing a nonlinear filter kernel.

6.4.2 Nonlinear NLMS

There is another way to regress bilinear or $\mathcal{O}(2)$ noise from a system signal; we feed in some pre-coupled data and pretend as if it is a genuine linear input. For example, to regress non-stationary bilinear noise with an LMS filter we can use the following as input

$$\mathbf{x}_{(1,2)}[n] = \begin{pmatrix} x_1[n]x_2[n] \\ x_1[n]x_2[n-1] \\ \vdots \\ x_1[n]x_2[n-N] \\ \vdots \\ x_1[n-N+1]x_2[n-N] \\ x_1[n-N]x_2[n-N] \end{pmatrix} \tag{6.32}$$

This would result in the coefficient update equation

$$\mathbf{a}_{(1,2)}[n+1] = \mathbf{a}_{(1,2)}[n] - 2\mu e[n]\mathbf{x}_{(1,2)}[n] \tag{6.33}$$

To make sure that loud transients do not change the filter coefficients for many iterations down the line, we can introduce a "leak" coefficient, α. Additionally, we can normalize the step-size by the power in the M-tap chunk of data we are considering. This effectively mitigates the slow converge due to a large difference in eigenvalues (consequence of a large power difference in the spectrum). This leads to (including a stabilization factor ψ)

$$\mathbf{a}_{(1,2)}[n+1] = (1-\alpha)\mathbf{a}_{(1,2)}[n] - \frac{2\mu}{\mathbf{x}^T_{(1,2)}[n]\mathbf{x}_{(1,2)}[n] + \psi} e[n]\mathbf{x}_{(1,2)}[n] \tag{6.34}$$

We could apply this same logic to a Wiener filter. That is, we assume some coupling between channels and feed this coupled channel into the filter as if it is a single channel and we attempt to remove the "linear" noise as represented by the input channel. This obviously requires nothing more than a Wiener filter or a single-channel LMS filter (and is consequently very fast and lightweight), so we then ask why should we go through the trouble of calculating the full second order Volterra kernel?

The reason is that there is only one set of weights for a linear filter. Consider an M-tap Wiener filter into which we are feeding a direct product of channels, $x[n] \equiv x_1[n] \cdot x_2[n]$. We will end up with M coefficients which encode how $x[n]$ couples into the system signal. This means that the contribution of $x_1[n]$ and $x_2[n]$ must be the same within the system. This is not necessarily true however. Contrast this with the coefficient matrix for the Volterra filter. Here, each channel can change independently of the other or be coupled in various, changing ways. In other words, we would have to either know ahead of time what the coupling of the noise is, or be very lucky in order to regress non-linear noise with a linear filter and nonlinear input.

Using equation (6.34) can still benefit the data quality efforts since any subtraction, however imperfect, can provide useful information about the nature of the noise coupling into the system. This can then be followed up on in more detail and with other types of analytic or machine learning techniques.

If we *do* know the form of the noise, then this can be a good way to go. Consider the modulation of the power lines (@ 60 Hz) due to the ASC noise (@ 0.5 Hz) resulting in power at 60 ± 0.5 Hz. Let us call the channels $x_{\texttt{ASC}}$ and $x_{\texttt{MAINS}}$. We can then take the product of the channels giving $x = x_{\texttt{ASC}} \cdot x_{\texttt{MAINS}}$ and then bandpass the input x. This will remove the side bands around the power line quickly and safely (if this is done in the time domain it may require a phase shift). An ordinary linear filter will then remove the 60 Hz line itself. In this way, we can in principle demodulate any channels which allow it.

While we cannot recover the non-linear-input-to-a-linear-filter results directly from

the complete second order filter, we can determine where the differences between the two reside. First, from Equation 18, we note that there will only two terms to cross-correlate instead of four. Because of this, we must have that $m = m'$ and $i = j$. Setting $x_1[n] \cdot x_2[n] = x_{12}[n]$, then Equation 18 reduces to

$$\left\langle d[k]x_{12}[k-j] \right\rangle = \sum_{m=0}^{M-1} a_k[m] \left\langle x_{12}[k-m]x_{12}[k-j] \right\rangle \tag{6.35}$$

Although the form and procedures are similar, the mathematical outcomes are not. In the full second order result, we cross-correlate the individual witness channels and then multiply the resulting matrices. In the non-linear input case the situation is reversed: we multiply the witness channels and cross-correlate the results. And since the product of cross-correlations is not the same as the cross-correlation of a product in general, these filters behave rather differently under most circumstances.

6.5 Concluding Remarks

Procedures for mitigating noise and transient artifacts without distorting or degrading the underlying signals are of paramount importance at LIGO and elsewhere. Beginning with the success of the linear and stationary Wiener filter, we adapted this filter to create adaptive filters able to remove non-stationary noise without being adversely affected by glitches. The analysis was extended to second order with the development of the Volterra filter that was able to remove the 60 Hz sidebands from the O3 Hanford dataset. The ongoing development of these analytic filter techniques is a critical one. As the detector noise floor is pushed down through hardware upgrades, more and increasingly complicated noise couplings will continue to appear and a filter suite will be a mandatory toolkit. Software analyses will be required to search though the large space of auxiliary channels and identify promising couplings and channel lists. One such method is a brute force, parallelized grid search using lightweight analytic filters to identify couplings of a specific order. The resulting channels can be fed into refined analytic or network-based data cleaning routines as described in chapter 5.

Chapter 7

Conclusion and Discussion

There have been many subjects described throughout the course of this thesis detailing gravitational wave (GW) astronomy and the Advanced LIGO (aLIGO) GW detectors. The overarching themes broadly fall into two categories: data quality and isotropic gravitational wave background (GWB) analysis.

Before we can enhance data quality, the behavior of the interferometers (IFOs) needs to be thoroughly understood. In chapter 2 we first showed that an experiment capable of measuring a strain of 10^{-22} was indeed possible. The IFOs were then dissected categorically by noise source as we recovered the majority of the noise floor of aLIGO as a superposition of linearly coupled, terrestrial noise sources. There remains unexplained noises in the $\sim 10 - 50$ Hz frequency band and I constructed and built upon parallelized coherence methods to monitor the low frequency coupling, including fully integrated user interfaces, interactive plots and hash tables for quick analysis.

Chapters 3 and 4 described the isotropic stochastic gravitational wave background (SGWB), current search pipelines, results, and foreground signal subtraction in future detectors such as Cosmic Explorer. In the present analysis, aLIGO had completed its third observing run and these results were used in conjunction with the results of the first two runs. The best upper limits for the SGWB place the normalized astrophysical background energy density, $\Omega_{GW}^{\alpha=2/3}(f)$, roughly an order of magnitude below the detectable limit. Since the SNR grows as $\sqrt{T}$ and more detector upgrades are anticipated, the SGWB search efforts will soon be able to claim statistically significant measurements of the background. With the sensitivity improved by $\sim 10x$ in Cosmic Explorer,

measurements of the background will in some sense be more difficult due to the large number of loud CBCs in the foreground. To address this, we developed a method which is able to create a mask of the inspiral events and remove them from the data thereby enhancing the visibility of the GW background. Since the CBC signals (the binary neutron stars in particular) spend many minutes in the detector band at low frequencies, this notching method removed the majority of the low frequency band and with it, much of the sensitivity of the analysis. Nevertheless, this novel approach puts a new upper limit on the SGWB for 3G detectors and serves as a benchmark for future work.

In chapters 5 and 6 we began developing techniques for removing noise coupled into the detector's output strain data. The simplest filter is the linear, first order, stationary filter also known as the Wiener filter. We adapted this filter to produce LMS and RLS filter variants and showed that these filters can be used to remove non-stationary noise while handling the transient glitches that are common to LIGO data. Expanding the filter kernel to second order, we then developed the nonlinear Volterra filter. Using this, we were able to subtract the bilinear 60 Hz sidebands from the Hanford O3 data in a lightweight and analytic way. These filters can be used not only to remove noise, but to brute force search through enormous list of auxiliary channels to find witnesses which may be useful in subtracting higher order noise. In an effort to go beyond second order or non-stationary filters, I constructed a convolutional neural network pipeline called `DeepClean` which is a fully generalized filter method able to capture linear, nonlinear and non-stationary noise sources. I have also shown that this method is able to safely preserve the CBC events that occur within the data. This work is continuing to be developed and is preparing to be run "online" in low-latency for aLIGO's fourth observing run and beyond. As it continues to run and characterize the data, `DeepClean` may become the first line of defence for mitigating noise at LIGO.

References

[1] B. P. Abbott, R. Abbott, T. D. Abbott, et al. Observation of gravitational waves from a binary black hole merger. *Physical Review Letters*, 116(6), Feb 2016.

[2] B. P. Abbott, R. Abbott, T. D. Abbott, et al. Gwtc-1: A gravitational-wave transient catalog of compact binary mergers observed by ligo and virgo during the first and second observing runs. *Physical Review X*, 9(3), Sep 2019.

[3] R. Abbott, T. D. Abbott, S. Abraham, et al. Gwtc-2: Compact binary coalescences observed by ligo and virgo during the first half of the third observing run, 2020, 2010.14527.

[4] B. P. Abbott, R. Abbott, T. D. Abbott, et al. Gw170817: Observation of gravitational waves from a binary neutron star inspiral. *Phys. Rev. Lett.*, 119:161101, Oct 2017.

[5] A. Einstein. *Sitzungsberichte der Koniglich Preußischn Akademie Wissenschaften (Berlin)*, volume 884-847. 1915.

[6] Michele Maggiore. Gravitational wave experiments and early universe cosmology. *Physics Reports*, 331(6):283–367, Jul 2000.

[7] B. P. Abbott, R. Abbott, T. D. Abbott, et al. A gravitational-wave standard siren measurement of the hubble constant. *Nature*, 551(7678):85–88, Oct 2017.

[8] J. H. Taylor and J. M. Weisberg. A new test of general relativity - Gravitational radiation and the binary pulsar PSR 1913+16. *Astrophysical Journal Letters*, 253:908–920, February 1982.

[9] R. A. Hulse and J. H. Taylor. Discovery of a pulsar in a binary system. *Astrophysical Journal Letters*, 195:L51–L53, January 1975.

[10] Rainer Weiss. Electromagnetically coupled broadband gravitational antenna. In *K.S. Thorne, "Gravitational radiation", 300 Years of Gravitation, S W Hawking and W Israel, pp 330–458*. University Press, 1972.

[11] A. Einstein. On the electrodynamics of moving bodies.

[12] Planck Collaboration, N. Aghanim, Y. Akrami, et al. Planck 2018 results. vi. cosmological parameters, 2020, 1807.06209.

[13] A. Buikema, C. Cahillane, G. L. Mansell, et al. Sensitivity and performance of the advanced ligo detectors in the third observing run. *Phys. Rev. D*, 102:062003, Sep 2020.

[14] The LIGO Scientific Collaboration, the Virgo Collaboration, R. Abbott, et al. Population properties of compact objects from the second ligo-virgo gravitational-wave transient catalog, 2020, 2010.14533.

[15] R. Abbott, T. D. Abbott, S. Abraham, et al. Gw190814: Gravitational waves from the coalescence of a 23 solar mass black hole with a 2.6 solar mass compact object. *The Astrophysical Journal*, 896(2):L44, Jun 2020.

[16] W. G. Anderson and D. E. Creighton. *Gravitational Wave Physics and Astronomy*. John Wiley & Sons, Ltd, 2011.

[17] Samantha A. Usman et al. The PyCBC search for gravitational waves from compact binary coalescence. *Class. Quant. Grav.*, 33(21):215004, 2016, 1508.02357.

[18] S. Chandrasekhar and E. Fermi. Problems of gravitational stability in the presence of a magnetic field. *The Astrophysical Journal*, 118:116, July 1953.

[19] M. Maggiore. *Gravitational Waves: Volume 1 - Theory and Experiments*. Oxford University Press, 2008.

[20] J. Aasi, J. Abadie, B. P. Abbott, et al. Gravitational waves from known pulsars: Results from the initial detector era. *The Astrophysical Journal*, 785(2):119, Apr 2014.

[21] M. Maggiore. *Gravitational Waves: Volume 2 - Astrophysics and Cosmology*. Oxford University Press, 2018.

[22] The LIGO Scientific Collaboration, the Virgo Collaboration, the KAGRA Collaboration, et al. Upper limits on the isotropic gravitational-wave background from advanced ligo's and advanced virgo's third observing run, 2021, 2101.12130.

[23] A. Buikema, C. Cahillane, G. L. Mansell, et al. Sensitivity and performance of the advanced ligo detectors in the third observing run. *Physical Review D*, 102(6), Sep 2020.

[24] B P Abbott, R Abbott, R Adhikari, et al. Ligo: the laser interferometer gravitational-wave observatory. *Reports on Progress in Physics*, 72(7):076901, Jun 2009.

[25] Chris Pankow, Katerina Chatziioannou, Eve A. Chase, et al. Mitigation of the instrumental noise transient in gravitational-wave data surrounding gw170817. *Physical Review D*, 98(8), Oct 2018.

[26] C. Biwer, D. Barker, J. C. Batch, et al. Validating gravitational-wave detections: The advanced ligo hardware injection system. *Physical Review D*, 95(6), Mar 2017.

[27] P. Nguyen, Schofield R. M. S., A. Effler, and C. Austin. Environmental noise in advanced ligo detectors. 2021.

[28] P. J. Sutton. S3 performance of the ligo interferometers as measured by sensemonitor. 2003.

[29] Lee Samuel Finn and David F. Chernoff. Observing binary inspiral in gravitational radiation: One interferometer. *Physical Review D*, 47(6):2198–2219, Mar 1993.

[30] B. P. Abbott, R. Abbott, T. D. Abbott, et al. Gw170814: A three-detector observation of gravitational waves from a binary black hole coalescence. *Physical Review Letters*, 119(14), Oct 2017.

[31] M. Tse, Haocun Yu, N. Kijbunchoo, et al. Quantum-enhanced advanced ligo detectors in the era of gravitational-wave astronomy. *Phys. Rev. Lett.*, 123:231107, Dec 2019.

[32] D. V. Martynov, E. D. Hall, B. P. Abbott, et al. Sensitivity of the advanced ligo detectors at the beginning of gravitational wave astronomy. *Physical Review D*, 93(11), Jun 2016.

[33] Giles Hammond, Stefan Hild, and Matthew Pitkin. Advanced technologies for future ground-based, laser-interferometric gravitational wave detectors. *Journal of Modern Optics*, 61(sup1):S10–S45, Jun 2014.

[34] F Matichard, B Lantz, R Mittleman, et al. Seismic isolation of advanced LIGO: Review of strategy, instrumentation and performance. *Classical and Quantum Gravity*, 32(18):185003, aug 2015.

[35] S. Gras and M. Evans. Direct measurement of coating thermal noise in optical resonators. *Physical Review D*, 98(12), Dec 2018.

[36] S. Ballmer et al. Thermal compensation system description. *LIGO Document Control Center, LIGO-T050064*, 2005.

[37] Rich Ormiston, Tri Nguyen, Michael Coughlin, et al. Noise reduction in gravitational-wave data via deep learning. *Physical Review Research*, 2(3), Jul 2020.

[38] Derek Davis, Thomas Massinger, Andrew Lundgren, et al. Improving the sensitivity of advanced ligo using noise subtraction. *Classical and Quantum Gravity*, 36(5):055011, Feb 2019.

[39] P. B. Covas, A. Effler, E. Goetz, et al. Identification and mitigation of narrow spectral artifacts that degrade searches for persistent gravitational waves in the first two observing runs of advanced ligo. *Physical Review D*, 97(8), Apr 2018.

[40] S. Soni. `Reduction and Identification of Scattered Light noise in LIGO`. https://dcc.ligo.org/LIGO-G2100105/public, 2021.

[41] S Soni, C Austin, A Effler, et al. Reducing scattered light in ligo's third observing run. *Classical and Quantum Gravity*, 38(2):025016, Jan 2021.

[42] Marissa Walker, Alfonso F Agnew, Jeffrey Bidler, et al. Identifying correlations between ligo's astronomical range and auxiliary sensors using lasso regression. *Classical and Quantum Gravity*, 35(22):225002, Oct 2018.

[43] P. M. Meyers and R. G. Ormiston. `STAMP-PEM`. https://git.ligo.org/patrick-meyers/stamp-pem/, 2017.

[44] `d3js` library. https://d3js.org/.

[45] Pablo A. Rosado. Gravitational wave background from binary systems. *Phys. Rev. D*, 84:084004, Oct 2011.

[46] Xing-Jiang Zhu, E. Howell, T. Regimbau, et al. Stochastic gravitational wave background from coalescing binary black holes. *The Astrophysical Journal*, 739(2):86, Sep 2011.

[47] S. Marassi, R. Schneider, G. Corvino, et al. Imprint of the merger and ring-down on the gravitational wave background from black hole binaries coalescence. *Phys. Rev. D*, 84:124037, Dec 2011.

[48] C. Wu, V. Mandic, and T. Regimbau. Accessibility of the gravitational-wave background due to binary coalescences to second and third generation gravitational-wave detectors. *Phys. Rev. D*, 85:104024, May 2012.

[49] Cheng-Jian Wu, Vuk Mandic, and Tania Regimbau. Accessibility of the stochastic gravitational wave background from magnetars to the interferometric gravitational wave detectors. *Phys. Rev. D*, 87:042002, Feb 2013.

[50] Pablo A. Rosado. Gravitational wave background from rotating neutron stars. *Phys. Rev. D*, 86:104007, Nov 2012.

[51] Alessandra Buonanno, Günter Sigl, Georg G. Raffelt, et al. Stochastic gravitational-wave background from cosmological supernovae. *Phys. Rev. D*, 72:084001, Oct 2005.

[52] Xing-Jiang Zhu, Eric Howell, and David Blair. Observational upper limits on the gravitational wave production of core collapse supernovae. *Monthly Notices of the Royal Astronomical Society: Letters*, 409:L132 – L136, 11 2010.

[53] B. P. Abbott, R. Abbott, T. D. Abbott, et al. Constraints on cosmic strings using data from the first advanced ligo observing run. *Phys. Rev. D*, 97:102002, May 2018.

[54] Xavier Siemens, Vuk Mandic, and Jolien Creighton. Gravitational-wave stochastic background from cosmic strings. *Phys. Rev. Lett.*, 98:111101, Mar 2007.

[55] Thibault Damour and Alexander Vilenkin. Gravitational radiation from cosmic (super)strings: Bursts, stochastic background, and observational windows. *Phys. Rev. D*, 71:063510, Mar 2005.

[56] Misao Sasaki, Teruaki Suyama, Takahiro Tanaka, and Shuichiro Yokoyama. Primordial black hole scenario for the gravitational-wave event gw150914. *Phys. Rev. Lett.*, 117:061101, Aug 2016.

[57] Vuk Mandic, Simeon Bird, and Ilias Cholis. Stochastic gravitational-wave background due to primordial binary black hole mergers. *Phys. Rev. Lett.*, 117:201102, Nov 2016.

[58] Sai Wang, Yi-Fan Wang, Qing-Guo Huang, and Tjonnie G. F. Li. Constraints on the primordial black hole abundance from the first advanced ligo observation run using the stochastic gravitational-wave background. *Phys. Rev. Lett.*, 120:191102, May 2018.

[59] Bruce Allen and Joseph D. Romano. Detecting a stochastic background of gravitational radiation: Signal processing strategies and sensitivities. *Physical Review D*, 59(10), Mar 1999.

[60] R. Abbott, T. D. Abbott, S. Abraham, et al. Observation of gravitational waves from two neutron star–black hole coalescences. *The Astrophysical Journal Letters*, 915(1):L5, Jun 2021.

[61] B. P. Abbott et al. GW150914: Implications for the stochastic gravitational wave background from binary black holes. *Phys. Rev. Lett.*, 116(13):131102, 2016, 1602.03847.

[62] E. S. Phinney. A Practical theorem on gravitational wave backgrounds. 7 2001, astro-ph/0108028.

[63] B. P. Abbott, R. Abbott, T. D. Abbott, et al. Upper limits on the stochastic gravitational-wave background from advanced ligo's first observing run. *Physical Review Letters*, 118(12), Mar 2017.

[64] Eric Thrane and Joseph D. Romano. Sensitivity curves for searches for gravitational-wave backgrounds. *Phys. Rev. D*, 88:124032, Dec 2013.

[65] B. P. Abbott, R. Abbott, T. D. Abbott, et al. Upper limits on the stochastic gravitational-wave background from advanced ligo's first observing run. *Physical Review Letters*, 118(12), Mar 2017.

[66] A. Matas, I. Dvorkin, T. Regimbau, and A. Romero. Applying gating to stochastic searches in o3. https://dcc.ligo.org/T2000512/, 2020.

[67] Kara M., Philippe N., Robert S., and Sharan B. Pem weekly magnetic injections. https://alog.ligo-wa.caltech.edu/aLOG/index.php?callRep=48212, 2019.

[68] Michael Coughlin, Alessio Cirone, Patrick Meyers, et al. Measurement and subtraction of schumann resonances at gravitational-wave interferometers. *Physical Review D*, 97, 02 2018.

[69] E. Thrane, N. Christensen, R. M. S. Schofield, and A. Effler. Correlated noise in networks of gravitational-wave detectors: Subtraction and mitigation. *Physical Review D*, 90(2), Jul 2014.

[70] Michael W Coughlin, Nelson L Christensen, Rosario De Rosa, et al. Subtraction of correlated noise in global networks of gravitational-wave interferometers. *Classical and Quantum Gravity*, 33(22):224003, Oct 2016.

[71] L. McCuller, C. Whittle, D. Ganapathy, et al. Frequency-dependent squeezing for advanced ligo. *Physical Review Letters*, 124(17), Apr 2020.

[72] David Reitze, Rana X Adhikari, Stefan Ballmer, et al. Cosmic explorer: The u.s. contribution to gravitational-wave astronomy beyond ligo, 2019, 1907.04833.

[73] Sheila Dwyer, Daniel Sigg, Stefan W. Ballmer, et al. Gravitational wave detector with cosmological reach. *Physical Review D*, 91(8), Apr 2015.

[74] B. O'Riley, M. Branchesi, S. Haino, and G. Gemme. Noise curves used for simulations in the update of the observing scenarios paper. https://dcc.ligo.org/LIGO-T2000012/public, 2020.

[75] M. Evans, J. Harms, and S. Vitale. Exploring the sensitivity of next generation gravitational wave detectors. https://dcc.ligo.org/LIGO-P1600143/public, 2016.

[76] M. Evans, S. Vitale, R. Sturani, and E. Hall. Unofficial sensitivity curves (asd) for aligo, kagra, virgo, voyager, cosmic explorer, and einstein telescope. https://dcc.ligo.org/LIGO-T1500293/public, 2020.

[77] Piero Madau and Mark Dickinson. Cosmic star-formation history. *Annual Review of Astronomy and Astrophysics*, 52(1):415–486, Aug 2014.

[78] Surabhi Sachdev, Tania Regimbau, and B. S. Sathyaprakash. Subtracting compact binary foreground sources to reveal primordial gravitational-wave backgrounds. *Physical Review D*, 102(2), Jul 2020.

[79] B. P. Abbott, R. Abbott, T. D. Abbott, et al. Binary black hole mergers in the first advanced ligo observing run. *Phys. Rev. X*, 6:041015, Oct 2016.

[80] B. P. Abbott, R. Abbott, T. D. Abbott, et al. Gw170104: Observation of a 50-solar-mass binary black hole coalescence at redshift 0.2. *Phys. Rev. Lett.*, 118:221101, Jun 2017.

[81] Elisabeth Vangioni, Keith A. Olive, Tanner Prestegard, et al. The impact of star formation and gamma-ray burst rates at high redshift on cosmic chemical evolution and reionization. *Monthly Notices of the Royal Astronomical Society*, 447(3):2575–2587, Jan 2015.

[82] Gregory Ashton, Moritz Hübner, Paul D. Lasky, et al. Bilby: A user-friendly bayesian inference library for gravitational-wave astronomy. *The Astrophysical Journal Supplement Series*, 241(2):27, Apr 2019.

[83] T. et al Regimbau. `MDC_Generation`. https://git.ligo.org/stochastic/MDC_Generation/, 2021.

[84] Patricia Schmidt, Ian W. Harry, and Harald P. Pfeiffer. Numerical relativity injection infrastructure, 2017, 1703.01076.

[85] Matteo Breschi, Rossella Gamba, and Sebastiano Bernuzzi. `bajes`: Bayesian inference of multimessenger astrophysical data, methods and application to gravitational-waves, 2021, 2102.00017.

[86] Rory J E Smith, Gregory Ashton, Avi Vajpeyi, and Colm Talbot. Massively parallel bayesian inference for transient gravitational-wave astronomy. *Monthly Notices of the Royal Astronomical Society*, 498(3):4492–4502, Aug 2020.

[87] Vijay Varma, Ryuichi Fujita, Ashok Choudhary, and Bala R. Iyer. Comparison of post-newtonian templates for extreme mass ratio inspirals. *Physical Review D*, 88(2), Jul 2013.

[88] Ashish Sharma and Jan Harms. Searching for cosmological gravitational-wave backgrounds with third-generation detectors in the presence of an astrophysical foreground. *Physical Review D*, 102(6), Sep 2020.

[89] Sylvia Biscoveanu, Colm Talbot, Eric Thrane, and Rory Smith. Measuring the primordial gravitational-wave background in the presence of astrophysical foregrounds. *Physical Review Letters*, 125(24), Dec 2020.

[90] Katarina Martinovic, Patrick M. Meyers, Mairi Sakellariadou, and Nelson Christensen. Simultaneous estimation of astrophysical and cosmological stochastic gravitational-wave backgrounds with terrestrial detectors. *Physical Review D*, 103(4), Feb 2021.

[91] Tanmaya Mishra, Brendan O'Brien, V. Gayathri, et al. Optimization of model independent gravitational wave search using machine learning, 2021, 2105.04739.

[92] Jingkai Yan, Mariam Avagyan, Robert E. Colgan, et al. Generalized approach to matched filtering using neural networks, 2021, 2104.03961.

[93] Grégory Baltus, Justin Janquart, Melissa Lopez, et al. Convolutional neural networks for the detection of the early inspiral of a gravitational-wave signal. *Physical Review D*, 103(10), May 2021.

[94] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus), 2020, 1606.08415.

[95] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014, 1412.6980.

[96] R. G. Ormiston. `CosmicExplore_MDC`. https://git.ligo.org/rich.ormiston/cosmicexplorer_mdc, 2021.

[97] R. Abbott, T. D. Abbott, S. Abraham, et al. Gw190521: A binary black hole merger with a total mass of 150 $M_{\odot}$. *Phys. Rev. Lett.*, 125:101102, Sep 2020.

[98] R. Abbott, T. D. Abbott, S. Abraham, et al. Gw190814: Gravitational waves from the coalescence of a 23 solar mass black hole with a 2.6 solar mass compact object. *The Astrophysical Journal*, 896(2).

[99] G. Vajente, Y. Huang, M. Isi, et al. Machine-learning nonstationary noise out of gravitational-wave detectors. *Phys. Rev. D*, 101:042003, Feb 2020.

[100] Ian J. Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, Cambridge, MA, USA, 2016. `http://www.deeplearningbook.org`.

[101] NVIDIA. Nvidia triton inference server, 2020.

[102] NVIDIA. Nvidia tensorrt, 2020.

[103] https://git.ligo.org/derek.davis/gwsubtract/blob/master/configs/h1-aux_options_c02.ini.

[104] J. C. Driggers, S. Vitale, A. P. Lundgren, et al. Improving astrophysical parameter estimation via offline noise subtraction for advanced ligo. *Physical Review D*, 99(4), Feb 2019.

[105] Sebastian Khan, Sascha Husa, Mark Hannam, et al. Frequency-domain gravitational waves from nonprecessing black-hole binaries. ii. a phenomenological model for the advanced detector era. *Physical Review D*, 93(4), Feb 2016.

[106] Mark Hannam, Patricia Schmidt, Alejandro Bohé, et al. Simple model of complete precessing black-hole-binary gravitational waveforms. *Physical Review Letters*, 113(15), Oct 2014.

[107] Sascha Husa, Sebastian Khan, Mark Hannam, et al. Frequency-domain gravitational waves from nonprecessing black-hole binaries. i. new numerical waveforms and anatomy of the signal. *Physical Review D*, 93(4), Feb 2016.

[108] Joshua S Speagle. dynesty: a dynamic nested sampling package for estimating bayesian posteriors and evidences. *Monthly Notices of the Royal Astronomical Society*, 493(3):3132–3158, Feb 2020.

[109] Ayon Biswas, Jess McIver, and Ashish Mahabal. New methods to assess and improve ligo detector duty cycle, 2019, 1910.12143.

[110] The LIGO Scientific Collaboration, the Virgo Collaboration, and the KAGRA Collaboration. Prospects for observing and localizing gravitational-wave transients with advanced ligo, advanced virgo and kagra, 2013, 1304.0670.

[111] V Tiwari, M Drago, V Frolov, et al. Regression of environmental noise in ligo data. *Classical and Quantum Gravity*, 32(16):165014, Jul 2015.

[112] E. Thrane, N. Christensen, R. M. S. Schofield, and A. Effler. Correlated noise in networks of gravitational-wave detectors: Subtraction and mitigation. *Physical Review D*, 90(2), Jul 2014.

[113] Nina Bode, Joseph Briggs, Xu Chen, et al. Advanced ligo laser systems for o3 and future observation runs. *Galaxies*, 8(4), 2020.

Appendix A

Analytic Filter Python Code Base

A.1 Extended Wiener Filter and Pipeline

```
def wiener_filter_pipeline(tar, wits, M):
    """
    run the wiener filter over the input data

    Parameters
    ----------
    M : `int`
        filter tap length
    tar : `numpy.ndarray`
        target array
    wit : `numpy.ndarray`
        stacked (or single) witness channel shape = [chans, timesteps]

    Returns
    -------
    clean : `numpy.ndarray`
        cleaned estimate of the true signal
    """
    W = extended_wf(tar, wits, M)
    est = np.zeros(tar.size)
```

```
    if wits.ndim == 1:
        wits = wits.reshape((1, wits.size))

    for ii in range(W.shape[0]):
        est += sig.lfilter(W[ii, :], 1.0, wits[ii, :])
    clean = tar - est
    return clean


def extended_wf(tar, wits, M):
    """
    Generalized wiener filter which takes advantage
    of the cross-correlations of the witness channels
    as well as the auto-correlations

    Parameters
    ----------
    tar : `numpy.ndarray`
        target signal
    wit : `numpy.ndarray`
        input signals (chans x lenght)
    M : `int`
        Filter order

    Returns
    -------
    W : `numpy.ndarray`
        filter coefficients (chans x coeffs)
    """
    if len(wits.shape) == 1:
        wits = np.reshape(wits, (1, wits.size))

    chans = wits.shape[0]
    combos = itertools.product(np.arange(chans), np.arange(chans))

    # cross correlation vector
```

```
    P = np.zeros((M+1) * chans)
    for ii in range(chans):
        p = cross_corr(tar, wits[ii, :], M)
        P[ii*(M+1):(ii+1)*(M+1)] = p[M:2*M+1]

    # correlation matrix
    R = np.zeros(((M+1)*chans, (M+1)*chans))
    for combo in combos:
        a, b = combo
        corr = cross_corr(wits[a, :], wits[b, :], M)
        toep = sl.toeplitz(np.flipud(corr[:(M+1)]), corr[M:2*M+1])
        R[a*(M+1):(a+1)*(M+1), b*(M+1):(b+1)*(M+1)] = toep

    W = np.linalg.solve(R, P)
    W = np.reshape(W, [chans, (M+1)])
    return W


def cross_corr(x, y, M=32):
    """
    channel cross correlation

    Parameters
    ----------
    x : `numpy.ndarray`
        channel 1
    y : `numpy.ndarray`
        channel 2
    M : `int`
        number of filter taps
    """
    xl = x.size
    xc = sig.fftconvolve(x, y[::-1])
    xc = xc[xl - 1 - M: xl + M]
    return xc
```

A.2 Leaky Normalized LMS Adaptive Filter

```
def nlms(d, x, M=1, mu=0.01, psi=0, leak=0):
    """
    Leaky, normalized LMS adaptive filter extended to
    analyze multiple inputs at a time

    Parameters
    ----------
    d : `numpy.ndarray`
        target array
    x : `numpy.ndarray`
        input channel
    M : `int`
        filter order
    mu : `float`
        step size
    psi : `float`
        stability factor when input power is low
    leak : `float`
        forgetting factor for weight updates

    Returns
    -------
    y : `numpy.ndarray`
        bilinear noise estimate
    """

    if len(x.shape) == 1:
        x = np.reshape(x, (1, x.size))

    estimate = np.zeros(x.shape[1])

    for ii in range(x.shape[0]):
        # initialize
        w = np.random.rand(M+1)
```

```
        e = np.zeros_like(d)
        y = np.zeros_like(d)
        xf = x[ii, :]

        if mu == None:
            eig = np.max(np.dot(xf, xf))
            mu = 1 / (4 * eig)

        # run the filter
        for k in range(M+1, d.size):
            xx = xf[k-(M+1):k]
            y[k] = np.dot(w, xx)
            e[k] = d[k] - y[k]
            w = w*(1-leak) + mu*e[k]*xx/(psi+np.dot(xx,xx))

        estimate += y

    return estimate
```

A.3 Volterra Filter and Pipeline

```
def volterra_pipeline(d, wit1, wit2, M):
    """
    Use this function to run the second order volterra filter
    in order to regress the dirty system signal d with the witness
    channels wit1 & wit2. Use M-taps. NOTE: the volterra kernel
    grows as (M+1)^4 so things will crawl to a stop very quickly if
    you're not careful!

    Parameters
    ----------
    d : `numpy.ndarray`
        noisy system signal
    wit1 : `numpy.ndarray`
        witness channel for noise regression
```

```
    wit2 : `numpy.ndarray`
        witness channel for noise regression
    M : `int`
        number of filter taps (a.k.a filter order)

    Returns
    -------
    clean : `numpy.ndarray`
        cleaned system signal
    """
    P = three_point_static_corr(d, wit1, wit2, M=M)
    out = four_point_corr(wit1, wit2, wit1, wit2, M)
    vc = out.reshape(((M+1)**2, (M+1)**2)).T
    weights = np.linalg.pinv(vc).dot(P)
    est = apply_weights_2d(wit1, wit2, weights.reshape(((M+1), (M+1))))
    clean = d - est
    return clean


def three_point_static_corr(a, b, c, M):
    """
    calculate the 3-point correlation of witness channels
    b & c with the target channel a. Mathematically, we find the
    matrix P by the following,

        P[i,j] = <a[k] b[k-i] c[k-j]>

    Note that the target array a does not get a time shift, hence
    the difference between a three-point correlation and a static
    three-point correlation

    Parameters
    ----------
    a : `numpy.ndarray`
        target array
    b : `numpy.ndarray`
```

```
        witness channe
    c : ‘numpy.ndarray‘
        witness channel
    M : ‘int‘
        number of filter taps (a.k.a filter order)

    Returns
    -------
    arr : ‘numpy.ndarray’
        2D bi-correlation array of size (M+1) x (M+1)
    """

    if a.size != b.size or b.size != c.size:
        sys.exit('Array sizes do not match!')

    # pad and stack (makes rolling cleaner)
    a = np.pad(a, (M,M))
    b = np.pad(b, (M,M))
    c = np.pad(c, (M,M))

    # fill the correlation matrix
    out = np.zeros((2*M+1, 2*M+1))
    for i in range(-M,M+1):
        for mp in range(-M,M+1):
            out[mp, i] = np.sum(a * np.roll(b, mp)\
                   * np.roll(c, i))

    out = out[:M+1, :M+1].flatten()
    return out


def four_point_corr(a, b, c, d, M):
    """
    calculate the 4-point correlation of witness channels
    a, b, c and d. Mathematically, we find the matrix C[m, m', i, j]
    by the following,
```

```
    C[m, m',i, j] = <a[k-m] b[k-m'] c[k-i] d[k-j]>

In practice, with the Volterra filter, a[k] = c[k] and
b[k] = d[k] giving

    C[m, m',i, j] = <a[k-m] b[k-m'] a[k-i] b[k-j]>

See the notes on overleaf here:
https://www.overleaf.com/project/5e2704c502c5d3000118937b

Parameters
----------
a : `numpy.ndarray`
    numpy array
b : `numpy.ndarray`
    numpy channe
c : `numpy.ndarray`
    numpy channel
c : `numpy.ndarray`
    numpy channel
M : `int`
    number of filter taps (a.k.a filter order)

Returns
-------
arr : `numpy.ndarray'
    2D bi-correlation array of size (M+1) x (M+1)
"""

if a.size != b.size or b.size != c.size or c.size != d.size:
    sys.exit('Array sizes do not match!')

# pad and stack (makes rolling cleaner)
a = np.pad(a, (M,M))
b = np.pad(b, (M,M))
```

```
    c = np.pad(c, (M,M))
    d = np.pad(d, (M,M))

    # fill the correlation matrix
    out = np.zeros((2*M+1,2*M+1, 2*M+1, 2*M+1))
    for j in range(-M,M+1):
        for i in range(-M,M+1):
            for mp in range(-M,M+1):
                for m in range(-M,M+1):
                    out[m, mp, i, j] = np.sum(np.roll(a, m) * np.roll(b, mp)\
                            * np.roll(c, i) * np.roll(d, j))

    out = out[:M+1, :M+1, :M+1, :M+1]
    return out


def apply_weights_2d(wit1, wit2, a):
    """
    Convolve 2d filter coefficients and the
    input witness channels.
    """
    M = a.shape[0] - 1

    wit1 = np.pad(wit1, (M, 0), constant_values=0)
    wit2 = np.pad(wit2, (M, 0), constant_values=0)

    y = np.zeros(wit1.size-M)
    for ii in range(wit1.size-M):
        wit_mat = np.outer(wit1[ii:ii + (M+1)][::-1], wit2[ii:ii+(M+1)][::-1])
        y[ii] = np.tensordot(wit_mat, a)

    return y
```